

ABSTRACT

Title of Document: THE ROLE OF EXECUTIVE FUNCTIONS IN
TYPICAL AND ATYPICAL
PRESCHOOLERS' SPEECH SOUND
DEVELOPMENT

Catherine Torrington Eaton, Doctor of
Philosophy, 2014

Directed By: Professor Nan Bernstein Ratner, Department of
Hearing and Speech Sciences

For most children, the acquisition of adult-like speech production is a seamless process. Yet for children with cognitive-linguistic speech sound disorder (SSD), in the absence of any obvious etiology such as hearing-related or motor processing deficits, the rules that govern their native phonology or speech sound system must be explicitly taught in speech therapy. A fundamental question asks why children with SSD are often unable to transition to adult-like production without direct therapy. One plausible, yet relatively unexplored explanation for this difficult transition is that there are differences in executive function abilities (EFs) in children with SSD as compared to typically-developing (TD) children. The core EFs (inhibitory control, cognitive flexibility, and working memory) are the cognitive functions needed to control initial or habituated impulses, shift flexibly between rule

sets, and store and manipulate information; these could logically be involved in the process of replacing early, inaccurate production patterns with adult phonology.

For this study, 4- to 5-year-old children, 20 with SSD and 45 with TD speech, participated in a battery of EF, speech production, and speech perception tasks. In addition, children were assessed using a modified version of the Syllable-Repetition Task (SRT; Shriberg et al., 2009), which is a variant of non-word repetition for children with SSD. Performance accuracy was compared across groups and also correlated with speech sound accuracy from a single-word naming task. It was found that children with SSD performed more poorly than the TD speech group on the forward digit span, SRT, and Flexible Item Selection (FIST; Jacques & Zelazo, 2001) tasks. Only forward digit span and SRT performances were positively correlated with speech production accuracy. Factor and regression analyses suggested that phonological memory capacity, but not inhibitory control, cognitive flexibility or mental manipulation is likely impaired in this population. Results from the SRT suggest that an additional cognitive component, such as phonological encoding or quality of underlying representations, may also be implicated. Interpretations for these and other results as well as their clinical implications are discussed.

THE ROLE OF EXECUTIVE FUNCTIONS IN TYPICAL AND ATYPICAL
PRESCHOOLERS' SPEECH SOUND DEVELOPMENT

By

Catherine Torrington Eaton

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:

Professor Nan Bernstein Ratner, Chair
Professor Rochelle Newman
Professor William Idsardi
Associate Professor Shelley Brundage
Assistant Professor Yi Ting Huang
Assistant Professor Donald Bolger

© Copyright by
Catherine Torrington Eaton
2014

Dedication

This project could not have been accomplished without the loving support of my family: Dave, Eva, Sayre, and my parents. Your sense of humor, cheerleading, and patience were a beacon for me throughout the entire process. Thank you for helping me keep life in perspective.

Acknowledgements

A heart-felt thank you to my mentors, Professors Nan Bernstein Ratner and Rochelle Newman, who provided me with intellectual guidance through each stage of this research. I am also grateful for the constructive feedback from my committee members, Drs Yi Ting Huang, Shelley Brundage, DJ Bolger and Bill Idsardi. I am sincerely appreciative of all the families who allowed their children to participate in this study as well as the clinicians and parents who provided me with referrals. I would also like to thank Lisa Freeman, Jenna Poland and the members of Nan's language lab for their invaluable assistance with coding, Monica Sampson for her scholarly advice and encouragement, and Yvan Rose for his assistance with Phon. Finally, I am grateful to funding from the University of Maryland's Graduate Dean's Dissertation fellowship, the Hearing and Speech Science department's MCM award, and a research assistantship from NSF grant BCS0745412.

Table of Contents

Dedication	ii
Acknowledgements	iii
Table of Contents	iv
List of Tables	v
List of Figures	vi
Chapter 1: Introduction	1
Chapter 2: Methods	28
Chapter 3: Results	65
Chapter 4: Discussion	92
Appendices	108
References	111

List of Tables

Table 1. Eligibility testing protocol and criteria

Table 2. Standardized test scores

Table 3. Demographic variables by group

Table 4. Statistical comparisons on standardized tests

Table 5. Statistical comparisons on spontaneous expressive language measures

Table 6. PCC-R for PNT by group

Table 7. EF task raw scores and statistics

Table 8. Partial correlation matrix between EF tasks and PNT

Table 9. Regression analyses of variables used to predict PNT scores

Table 10. Factor analysis of EF tasks with two-factor solution

Table 11. Modified SRT accuracy percentages and statistics by group

Table 12. Partial correlation matrix for mod-SRT

Table 13. Mean proportions of accurate consonants and between group statistics by stress pattern and stressed syllables

Table 14. Correlations between mod-SRT and EF tasks

Table 15. Regression analyses of mod-SRT and forward digit span as predictors of PNT scores

List of Figures

Figure 1. Flow chart of testing with numbers of participants

Figure 2. Order of experimental tasks and number of trials

Figure 3. DNS task example

Figure 4. Hearts and Flowers task design

Figure 5. FIST example

Figure 6. Animal span example

Figure 7. Results of the mod-SRT by group and task condition

Figure 8. Results of the mod-SRT by group and stimulus length

Chapter 1: Introduction

Overview

Although it takes years for phonological and articulation systems to mature, most children learn the sounds of their language without instruction. Phonology develops in individualized yet recognizable phases that are influenced by both genetics and environment (Bernthal, Bankson & Flipsen, 2009). By the age of 4, most children's speech is 90% intelligible to unfamiliar listeners and phonemes include all place and manner classes. By age 8, typically-developing children have developed adult-like phonological systems.

Yet for some children, in the absence of any obvious etiology such as hearing-related or motor deficits, the rules that govern their native speech sound system must be explicitly taught. Cognitive-linguistic speech sound disorder (SSD) is a developmental disorder of unknown etiology in which speech production is characterized by a high number of speech sound omissions, substitutions, and/or epenthesis that negatively impact a child's intelligibility. Fortunately, speech therapy, using a variety of approaches, has been found to be quite effective in remediating this disorder (Gierut, 1998; Law, Garrett & Nye, 2010).

A fundamental question that can be asked of children with SSD is why, unlike typically-developing (TD) children, they need phonological rules to be defined for them. While researchers have proposed deficits in speech perception (e.g., Rvachew, Rafaat & Martin, 1999), poorly-defined underlying representations (e.g., Sutherland & Gillon, 2007), and other theories (e.g., Munson, Edwards & Beckman, 2005), few have considered examining the domain-general cognitive mechanisms that are believed to

underlie language acquisition more generally. Cognitive flexibility, inhibitory control, and working memory, often considered the core executive functions (EFs; Diamond, 2013; Miyake, Friedman, Emerson, Witzki & Howerter, 2000), should logically be involved in the process of replacing early-developing, inaccurate and habituated production patterns with adult-like phonology. However, to date, very little research has investigated whether any, all, or a combination of these mechanisms might be implicated in this prevalent childhood communication disorder.

The purpose of this study is to examine whether impairments in core executive functions might in part explain why children with cognitive-linguistic SSD fail to spontaneously correct their disordered phonological productions. If impaired executive functioning is a contributing factor in the disorder, then children with phonological deficits should perform more poorly on core executive function tasks than age-matched peers with more mature profiles of articulation. Likewise, task performance on some or all of the EF domains should directly relate to speech sound accuracy. The next section overviews cognitive-linguistic SSD, presents some basic definitions of the core executive functions and what is known about the course of their development, reviews the limited available evidence on the relationship between executive functions and SSD, and discusses the theoretical rationale and research hypotheses proposed in this study.

Cognitive-linguistic speech sound disorder

Speech sound disorder is an umbrella term that encompasses a number of developmental speech production disorders of articulation, phonology, motor planning, and intelligibility. As a group, SSD is one of the most prevalent childhood

communication disabilities, affecting approximately 5% of children entering the first grade (National Institute on Deafness and Other Communication Disorders [NIDCD], 2006-2008). Many studies have demonstrated the negative and life-long impact of SSD on later academic performance and literacy (Felsenfeld & Broen, 1992; Lewis, Freebairn, & Taylor, 2000), psychosocial well-being (McCormack, McLeod, McAllister, & Harrison, 2010), and even future employment choices (McCormack, McLeod, McAllister & Harrison, 2009). In early childhood, SSD has also been implicated in slow vocabulary development and late talking, although it appears difficult to separate SSD and language disorders in this population (MacRoy-Higgins & Schwartz, 2013; Rescorla & Ratner, 1996; Schwartz & Leonard, 1982; Thal, Oroz & McCaw, 1995).

Researchers differ on how they subgroup speech sound disorders and thus terminology can be confusing (see Waring & Knight, 2013, for review). Bernthal, Bankson and Flipsen (2009) classify SSD into two main groups: 1) organically-based disorders associated with or resulting from genetic syndromes (e.g., Fragile X), neuromotor conditions (e.g., cerebral palsy and developmental apraxia of speech) and hearing loss, and 2) cognitive-linguistic disorders with no known origin. With regard to the latter group, which is also the most common, researchers have devised different systems of classification by possible etiology (Rvachew & Brosseau-Lapre, 2013; Shriberg et al., 2010), psycholinguistic deficit (Stackhouse & Wells, 1993), or symptomatology (Dodd, 2005). The subgroup of interest in this study is called cognitive-linguistic speech sound delay or disorder by Shriberg and colleagues, but it is more commonly known as phonological disorder in speech-language pathology. For purposes of this study, the clinical group will be referred as children with SSD.

Cognitive-linguistic SSD is characterized by an over-abundance of speech sound or syllable structure errors beyond what is expected given normative data on typical speech development. The disorder includes varying levels of severity, types of errors, and consistency of errors. Levels of severity range from very mild yet age-inappropriate articulatory errors, such as distorted production of /s/ or /r/, to more severe levels in which the percent of consonants produced correctly is less than 50% (Shriberg & Kwiatkowski, 1982). Researchers have distinguished errors of phoneme or syllable omission and substitution as either typically-developing or atypical errors (Preston & Edwards, 2010; Rvachew, Chiang & Evans, 2007). Typical errors include ‘phonological processes’ such as stopping of fricatives (“side” → “tide”), weak syllable deletion (“telephone” → “tefone”), and consonant cluster reduction (“star” → “tar”). Atypical errors, which can disproportionately affect intelligibility, include backing of frontal stops (“boy” → “goy”), substitution of fricatives with nasal air emission, and omission of initial singleton or consonant clusters (“star” → “ar”). Finally, differences in error consistency, not associated with developmental apraxia of speech, have been documented by a number of researchers (Dodd, 2005; Tyler, Lewis & Welsh, 2003).

The question of why children with cognitive-linguistic SSD fail to spontaneously correct their error patterns as typically-developing children do, has been a subject of heated debate ever since speech-pathology became a profession. Researchers have exhaustively investigated a wide array of potential underlying causes, from intelligence (Winitz, 1969), to oral-motor ability (see Powell, 2008), to haptic skills (Hardcastle, Gibbon, & Jones, 1991), and too many others to do justice to here. Most have not borne fruit, or have failed to describe the majority of children with cognitive-linguistic SSD. In

fact, most current textbooks do not devote much attention to potential etiological mechanisms at all (compare Bernthal et al., 2009 to Winitz, 1969). The single exception, which continues to be controversial, is the contention that children with cognitive-linguistic SSD have atypical speech perception skills.

Rvachew and others have proposed that cognitive-linguistic SSD stems from underlying deficits in speech perception that in turn adversely affect the development of phonemic representations and speech sound production (Rvachew & Brosseau-Lapre, 2013). A seminal study by Rvachew and Jamieson (1989) focused on discrimination of fricative contrasts and found that children in an articulation-disordered group performed worse than their same-aged peers and adults. Rvachew and colleagues have since published treatment studies comparing phonological treatment with and without perception-based training using a computerized speech discrimination protocol (Rvachew, Nowak & Cloutier, 2004; Rvachew et al., 1999). These combined results suggest that at least some children with cognitive-linguistic SSD present with differences in speech sound discrimination abilities.

One reason that the etiology of cognitive-linguistic SSD has proven difficult to disentangle is because of likely subgroups within the disorder. Some researchers debate whether there are discrete subgroups or a continuum along which deficits are either articulation-based or phonological. This distinction is often assessed through the practice of stimulability, or testing whether the child has the prompted ability to produce the sound (Lof, 1996; Miccio, Elbert & Forrest, 1999). Recent work in genetics proposes that identifications of subgroups may be related to heritability of the disorder. The search for reliable phenotypes to identify subgroups is ongoing (Shriberg et al., 2005; Stein et al.,

2010), though frequent co-morbidity of speech and language disorders makes the task challenging. Recent evidence also suggests possible neurobiological and/or neurophysiological differences in older children with persistent SSD (i.e., those continuing to produce speech sound errors beyond age eight), which may represent a distinct subgroup as well (Preston et al., 2014).

Fortunately, most children with cognitive-linguistic SSD respond favorably to treatment, evidencing gains in both accuracy and intelligibility (Gierut, 1998; Law et al., 2010). Interventions differ in how they teach the child. Some protocols delineate phonemic categories through contrasts in meaning (Gierut, 1991; Williams, 2003). Other protocols use pre-existing features in the child's sound system to establish new phonemes or word positions (Bernhardt & Stoel-Gammon, 1994). As previously mentioned, some protocols include a speech sound perception component (Rvachew et al., 2004). Still others teach the child to recognize for him or herself what needs to be changed (Dean & Howell, 1986; Hasketh, 2010).

Although studies comparing treatment approaches are relatively fewer in number (e.g., Hulterstam & Nettlebladt, 2002; Klein, 1996), most interventions are shown to be effective. The fact that children with cognitive-linguistic SSD respond to any number of interventions could suggest an underlying process that has yet to be explored. Though all young children produce simplified and often highly inaccurate versions of adult word forms during speech development, children with cognitive-linguistic SSD have difficulty overwriting their early productions without some form of explicit guidance. Domain-general cognitive processes might play an important role in the transition to adult-like phonology.

Core executive functions and preschoolers

According to Diamond and Lee (2011), “Executive Functions (EFs) are the cognitive control functions needed when you have to concentrate and think, when acting on our initial impulses might be ill-advised” (pp. 959). The three core executive functions are considered to be inhibitory control, cognitive flexibility, and working memory (Diamond, 2013). Over the past two decades, researchers have gained a greater understanding of the role that these EFs play in language and higher cognitive functions (e.g., Diamond, 2013), of the different subcomponents that are involved (Miyake et al., 2000) and of some of the neural bases of these cognitive mechanisms (e.g., Cocchi, Zalesky, Fornito & Mattingley, 2013; Diamond, 2011). The purpose of this section is not to provide a comprehensive review of the research behind these broad constructs, but rather to offer conceptual definitions of the three EFs of interest, describe how these constructs have been proposed to operate in adults and children, and review what is known of their development during the preschool years.

Operational definitions of the core EFs. Inhibitory control is widely recognized as consisting of two processes: the ability to ignore distracting information and the ability to stop an inappropriate response (Simpson & Riggs, 2007). The ability to ignore irrelevant information is less significant to the research hypotheses of this study, but the ability to suppress inappropriate responses could be quite relevant. This latter process is part of the larger notion of self-control, which also involves the concepts of self-discipline, controlling impulsivity, and delaying self-gratification (Diamond, 2013). Inhibitory control is often assessed through Go/No-Go or Stroop paradigms in which participants must actively inhibit prepotent verbal or motor responses under certain task-

defined conditions. In terms of clinical significance, inhibitory control has been implicated most commonly in attention deficit hyperactivity disorder (Schoemaker et al., 2012) and psycho-social disorders (Aksan & Kochanska, 2004).

Cognitive flexibility is defined as the ability to provide an appropriate response and then, when the task changes, to quickly shift to a different, also appropriate response (Deak, Ray & Pick, 2004). The term cognitive flexibility is used somewhat interchangeably with task-switching or set-shifting, although it also can be seen as a process that underlies those abilities (Diamond, 2013). Tasks that are used to assess this construct, such as card sorting tasks (Stuss, Levine, Alexander, Hong, Palumbo, et al., 2000; Zelazo, Frye & Rapus, 1996) typically require participants to provide a verbal or motor response to one salient feature and then switch to a second salient feature within an experimental block, inducing effects known as local switch costs (Koch, Gade, Schuch & Philipp, 2010). Switch rather than non-switch trials are typically used as the more informative measure for accuracy and reaction time (Davidson, Amso, Cruess Anderson & Diamond, 2006; Deak et al., 2004). Results from clinical populations such as individuals with autism spectrum disorders and traumatic brain injury have demonstrated cognitive inflexibility with certain tasks in comparison to performance of typical populations (Dockree & Robertson, 2011; Yerys, Wolff, Moody, Pennington & Hepburn, 2012).

Finally, working memory is defined as the ability to temporarily store information and mentally manipulate it (Davidson et al., 2006). In the context of core executive functions, working memory is also referred to as updating (Dauvier, Chevalier & Blaye, 2012; St Clair-Thompson, 2011), which highlights its role in the performance of goal-

directed tasks. Researchers agree that working memory is comprised of two discrete processing systems- one for verbal information, the area of interest for this study, and one for visuo-spatial information- and that these two systems are separate in young children as well (Alloway, Gathercole & Pickering, 2006; Baddeley, 2001).

According to Baddeley's model (2001; Repovs & Baddeley, 2006), temporary storage of speech information is maintained in the phonological loop, whereas the central executive accomplishes manipulation of information in addition to other functions (e.g., allocation of attentional resources). The phonological loop includes both storage and subvocalic rehearsal, which helps counteract the effects of decay. Storage capacity is most often assessed through serial recall of digits, words, or non-words. Since the phonological loop is seen as a slave system to the central executive, cognitive manipulation cannot be tested independently of capacity. Tasks such as re-ordering of items in memory have been used to assess the manipulation component (Diamond, 2013). Though working memory deficits have been associated with a number of clinical conditions, children with specific language impairment (SLI) provide some of the most compelling evidence with regard to verbal working memory impairments (Coady & Evan, 2008; Gathercole, 2006).

EF development in typically-developing children. Much of what is known about inhibitory control, cognitive flexibility, and phonological working memory comes from the literature on adult populations (e.g., Diamond, 2013, for review; Miyake et al., 2000). The three core EF constructs are often recruited in the same tasks and are seen as mutually supportive (Davidson et al., 2006; Diamond, 2013; Marcovitch & Zelazo, 2009). Neurocognitive and behavioral research with adults, however, suggests that they

can also be identified as distinct processes (Badre, 2008; Chien & Fiez, 2001; Frank, 2006; Miyake et al., 2000; Østby, Tamnes, Fjell & Walhovd, 2011).

Studies with EFs in young children are fewer in number, partially because of the language comprehension and attention required in testing paradigms, although the last decade has seen a burgeoning of behavioral methodologies and analytical techniques that have greatly informed our understanding. Although a few studies have tested two-year-olds (e.g., Carlson, 2005; O'Sullivan, Mitchell & Daehler, 2001), the majority of research examining core EFs begins at age three or four. The development of EFs is largely driven by neurobiological changes and learning. Some of these changes include increases in gray matter volume in pre-frontal regions, greater connectivity between regions, and fine-tuning of structures engaged in relevant tasks (Amso & Casey, 2005; Østby et al., 2011).

Although the exact neurobiological mechanisms underlying EFs in children have yet to be identified, behavioral studies suggest that rapid changes in the system occur between ages three and six. For instance, research has shown that three-year-olds have difficulty switching task rules and inhibiting prepotent or perseverative responses even when they are able to verbalize what is being asked of them (Zelazo, 2004), while a majority of four-year-olds succeed in these tasks (Carlson, 2005). Further evidence suggests that there are large differences in all EF abilities between the ages of three and four-and-a-half (Carlson, 2005; Diamond, Kirkham & Amso, 2002). Until the neural architecture is in place to support these cognitive processes, some researchers have observed that young children at different ages employ alternate strategies to accomplish tasks that older children or adults accomplish more directly (Davidson et al, 2006; Dauveir et al, 2012; Ramscar, Dye, Gustafson & Klein, 2013).

The underlying organization of and relationship between EF components in preschool children is the subject of considerable debate in the literature (e.g., Chevalier et al., 2012; St Clair-Thompson, 2011). Factor and confirmatory factor analyses have been used to measure performance of typically-developing children ages two to six on a myriad of tasks. These statistical methodologies compare models of best fit, whereby tasks factor together according to latent variables. Results have supported several models of EF in preschoolers, including a unitary construct of executive control (Shing, Lindenberger, Diamond, Li & Davidson, 2010; Wiebe, Andrews Espy & Charak, 2008), a system similar to that observed in adults in which all three core EF components can be isolated (Garon, Bryson & Smith, 2008; Diamond, 2013), and a two-construct system with only inhibitory control and working memory (Miller, Giesbrecht, Muller, McInerney & Kerns, 2012).

Further confounding the understanding of EF organization is evidence showing that the manipulation of tasks through external cues or prompts can facilitate preschoolers' success by scaffolding certain cognitive skills (Dauvier et al., 2012; Low & Simpson, 2012). The effects of these manipulations have been attributed to increases in conscious awareness, lessening of working memory demands, and overcoming attentional inertia (Deak et al., 2004; Diamond et al., 2002; Dowsett & Livesey, 2000; Kirkham, Cruess & Diamond, 2003; Marcovitch & Zelazo, 2009; Ramscar et al., 2013; Yerys & Munakata, 2006; Zelazo, 2004). These results also suggest the necessity of selecting appropriately sensitive tasks to test these constructs (Carlson, 2005).

In summary, adult models of inhibitory control, cognitive flexibility, and phonological working memory have been used to understand these processes in children.

However, because of rapidly changing neural underpinnings, particularly in the preschool years, it appears likely that children accomplish EF tasks differently at various stages of development, emphasizing the importance of age as a factor in performance. The exact relationship among core EFs in preschoolers is unknown, and appears greatly influenced by task demands.

Speech sound disorder and executive functions

Speech sound development can be seen as a process whereby early, simplified word forms are gradually replaced by production patterns that more closely match the child's input. Each of the three core EFs may play distinct roles in this process. Inhibitory control, for instance, could be required to suppress well-habituated word forms in lieu of correct productions. Cognitive flexibility could theoretically underlie a child's ability to overwrite early phonological rules or inaccurate representations in order to adopt the mature sound system. Working memory might be required to hold early templates in mind while manipulating and correcting speech sound targets. An impairment in one or a combination of core EFs might delay or hinder the transition to adult-like speech until more explicit instruction facilitates the process.

The role of executive functions in children with cognitive-linguistic SSD has been relatively unexplored. A handful of researchers have used well-established EF experimental paradigms with this population, while others have provided more indirect evidence of EF abilities. Empirical evidence is likely to inform both the relationship between speech and EF development as well as clinical theory and management. This

section is a review of the evidence to date regarding each core EF in this population; it ultimately highlights the need for additional research.

Evidence of inhibitory control deficits. Inhibitory control in children with cognitive-linguistic SSD has not been explored experimentally, but there is indirect evidence from studies of one therapeutic approach that suggests it may be an area of weakness. Research has shown that correcting sounds in error by training unfamiliar words or non-words facilitates learning (Cummings & Barlow, 2011; Gerber, 1973; Gierut, Morrisette, & Ziemer, 2010). Gierut et al. (2010) published results of a treatment study comparing two groups of children (N = 30) with cognitive-linguistic SSD. The groups were balanced on all aspects of speech and language, duration of therapy, and therapeutic approach; however, one group was trained using non-words while the other was treated using real words. Group gains were measured by accuracy of trained targets as well as accuracy of untrained, real word items with the participants' target sound or sounds. Results demonstrated that the group trained on non-words evidenced faster gains in learning of trained items as well as faster and better maintained gains on untrained real word targets than the control group.

The researchers highlight that this study and similar ones show efficiency of treatment in the absence of established semantic knowledge. Though the non-words in this study were treated as novel real words (i.e., nouns and verbs with meanings exemplified in pictures) rather than meaningless syllable strings, children were better able to correct errors without older, habituated productions interfering with learning. The results can be evaluated in terms of inhibitory control; perhaps children with cognitive-linguistic SSD fail to spontaneously correct their errors because of difficulty inhibiting

inaccurate, prepotent word forms. Teaching novel words may help bridge the transition to correct phonology.

Evidence of cognitive flexibility deficits. A source of evidence with regard to cognitive flexibility in this population uses more established EF paradigms to suggest an area of weakness. Dodd and her colleagues (Crosbie, Holm & Dodd, 2009; Dodd, 2011; Dodd & McIntosh, 2008) have argued that children with cognitive-linguistic SSD can be subdivided by symptomatology, which ultimately reflects differences in underlying etiology. They suggest that children from one specific subtype of cognitive-linguistic SSD- those with consistent but atypical speech sound errors- perform more poorly on cognitive flexibility tasks than children from other subtypes of SSD and TD children.

Dodd and her colleagues have used two paradigms to test their hypothesis (Crosbie et al., 2009; Dodd & McIntosh, 2008). In one task, children participated in a computer program whereby selecting a particular shape and/or color resulted in a rewarding visual display. Once the participant learned to select the appropriate response item, the criterion rule changed. Participants were tested on how quickly (out of a possible four chances) they were able to abstract the new rule. The second measure of executive functioning was examined using the Flexible Item Selection Test (FIST; Jacques & Zelazo, 2001). In this paradigm, children were asked to identify two related cards on a particular dimension (color, size, shape) out of a set of three. After choosing the first pair, participants were asked to select two cards out of the same set that were related on a different dimension. The groups showed significant differences on both tasks. Children with consistent, atypical errors abstracted fewer rules on the computer

task than other speech-impaired and TD groups and were less successful on both first and second selections of the FIST.

Dodd and her colleagues' interpretation of these findings is that children with a particular subtype of cognitive-linguistic SSD have deficits in the ability to abstract rules- such as the rules governing native-language phonology- rather than an impairment of cognitive flexibility *per se*. However, there are several aspects of their research that make it difficult to interpret under a framework of executive functioning. First, they included both the first and second selections in the FIST (i.e., choosing one pair on a given attribute and then a second pair on a different attribute). This method of coding differs from what is used as a measure of cognitive flexibility (Jacques & Zelazo, 2001), in which the second choice contingent on an accurate first choice is the target response where children must shift their thinking. In addition, methodological details such as participant characteristics and phonological scoring criteria are relatively sparse in these reports. Finally, assessment of inhibition and working memory, both or either of which could conceivably affect task performance, are not included in these studies. Nonetheless, Dodd and colleagues' research suggests the value of follow-up study.

Evidence of working memory deficits. Although working memory has been strongly implicated as an underlying deficit in children with SLI (e.g., Briscoe & Rankin, 2009; Coady & Evans, 2008), it has not frequently been examined in children with SSD without concurrent language impairment. Many studies involving preschool children with SLI either choose not to measure speech sound accuracy or fail to factor in its likely influence on performance (Adams & Gathercole, 2000; Archibald & Gathercole, 2007).

Thus, there is a possibility that participants in these studies include children with co-occurring SLI and SSD.

Several studies have sought to examine phonological memory in children with relatively pure cognitive-linguistic SSD compared with other subtypes of SSD, co-occurring SSD with SLI and/or reading impairment, and TD children (Farquharson Schussler, 2013; Lewis et al., 2006; Lewis et al., 2011; Shriberg et al., 2009). A few of these studies have used the forward digit span task to assess phonological storage capacity. For instance, Tkach et al. (2011) conducted a neuroimaging study comparing six children with history of cognitive-linguistic SSD and children with typical speech. Behavioral results showed that all members of the clinical group scored below average on forward digit span. Likewise, the results of a heritability study by Lewis et al. (2011) reported lower performance on forward digit span by probands and affected siblings as compared to unaffected siblings, although the differences were not statistically significant.

Considerably more evidence for this EF construct in cognitive-linguistic SSD is found using the paradigm of non-word repetition (NWR). This task has been considered a valid measure of phonological memory that is relatively free from the influence of lexical knowledge and well-habituated word forms (Gathercole, 2006). One disadvantage that is frequently noted in the NWR literature is that in addition to phonological memory, the task likely involves a number of other levels of processing including perceptual/auditory analysis, phonological encoding, phonological and articulatory planning, and motor output (Coady & Evans, 2008; Graf Estes, Evans & Else-Quest, 2007). With regard to the level of planning and motor output, researchers who have used this paradigm with

children with SSD have adjusted their scoring in various ways to accommodate consistent sounds in error that otherwise confound overall accuracy, which will be discussed further in the next section.

Results from these studies have consistently demonstrated that children with SSD have lower NWR accuracy than children with typical speech, even when articulation is controlled. Furthermore, as length increases, accuracy decreases, which is a trend known as the length effect, a finding demonstrated across populations (Archibald & Gathercole, 2007; Lewis et al., 2011; Munson et al., 2005; Preston & Edwards, 2007; Roy & Chiat, 2004; Shriberg et al., 2009; Sutherland & Gillon, 2007). Although many of these researchers have interpreted group differences in NWR task performance as an indication of weak phonological representations, these results also suggest a deficit in phonological memory.

One finding that has been raised as a concern by Shriberg et al. (2009), is that children with SSD performed more poorly than TD children on two-syllable items rather than solely on multi-syllabic stimuli. Arguably the shortest items should not significantly tax memory and therefore should be produced with similar accuracy by both groups of participants. An additional error analysis led the researchers to suggest that children with cognitive-linguistic SSD also have deficits in auditory-perceptual encoding. These and similar data have led other researchers to question whether short-term storage is the primary cognitive process that is measured in this task (Graf Estes et al., 2007).

Summary. The evidence covered thus far is only suggestive of differences in core EFs in children with cognitive-linguistic SSD. In general, research to date lacks the experimental control needed to answer questions regarding the relationship between EFs

and phonological development in the preschool years (Dodd, 2011; Gierut et al., 2010; Tkach et al., 2011). In addition, the relationship between NWR and other working memory tasks is not entirely straightforward. NWR likely includes processes that other working memory tasks do not engage (Archibald & Gathercole, 2007; Graf Estes et al., 2007; Shriberg et al., 2009). Conversely, NWR lacks a cognitive manipulation component, the second process included in the operational definition of working memory (Diamond, 2013). A study that is explicitly designed under an EF framework to explore the relationships between EF tasks, NWR, and speech sound accuracy may help explain the difficulties faced by children with cognitive-linguistic SSD.

Non-word repetition

Controlling articulation/phonological artifacts. As discussed previously, NWR has been proposed to be a fairly reliable measure of phonological memory (Gathercole, 2006) and a relatively sensitive clinical marker of SLI (e.g., Briscoe & Rankin, 2009). If the goal of this task is to provide a valid measure of phonological or working memory storage, the NWR task should control for processing levels that are known to be impaired or under-developed in specific populations or age groups. For purposes of this study, phonological and articulation output must be controlled in children with cognitive-linguistic SSD to prevent significantly confounding results. That is, non-word repetition performance will be conflated by consistent speech production errors if not effectively controlled.

Researchers have proposed various methods of controlling variability from phonological or articulatory immaturity or impairment. Stokes and Klee (2009), for

example, adapted their procedures with toddlers by scoring consistent sounds in error as correct. Hoff, Core, and Bridges (2008) statistically controlled production of non-words by partialling out real word accuracy from a list of phonemically similar words. Roy and Chiat (2004) pre-specified a number of common phonological variants or errors that would be accepted across participants. Dollaghan and Campbell (1998) established a new protocol called the Non-word Repetition Task (NRT) that purposely excluded later developing phonemes. These are just a few examples of some of the methodologies used to remove the unintended effects of articulation from task performance.

Recently, Shriberg and colleagues (2009) proposed an alternate non-word repetition task known as the Syllable Repetition Task (SRT). The SRT consists of only four early-developing consonants and one vowel. In addition, items are produced as syllable strings without overt lexical stress. The task was normed on 158 three- to five-year-old children, 95 of whom had speech and/or language delay. Some of their significant findings included: 1) a high correlation between performance on the SRT and the NRT (Dollaghan & Campbell, 1998), 2) a high correspondence between the five phonemes used in the SRT and those same phonemes in the participants' speech sound inventories, and 3) good specificity and sensitivity in identifying children with language impairment as compared to the NRT. As noted earlier, the results also demonstrated poorer performance on both the NRT and SRT by children with cognitive-linguistic SSD (and typical expressive language) as compared to TD children. Taken together, the SRT appears to be a psychometrically sound test for this clinical population that effectively controls for articulatory confounds. As with any new methodology, replication is needed to support its validity.

Lexical stress in NWR. One aspect of the SRT's design that could be manipulated is the addition of prosody. Although word-level stress has been found to be independent of segmental encoding (Biran & Friedmann, 2006; Marton, 2006), it likely plays an important role in phonological memory. First, word-level stress promotes the process of 'redintegration' or using lexical activation to support representations stored in the phonological loop (Gathercole, 2006). Studies have demonstrated that non-words that are more word-like are more easily recalled because of this process (Gathercole, 1995). Second, because of the small phoneme inventory and equal-stressed, monotone presentation of the SRT, items are highly redundant, which could negatively affect performance due to interference among stimuli. Third, word-level stress promotes the mnemonic strategy of chunking phonemic information, which has been found to aid short-term storage (Chen & Cowan, 2005). For these reasons, it might be expected that performance on the SRT would be worse than on typical non-word paradigms because SRT items are perceived as meaningless syllable strings rather than possible lexical items.

Removing prosody from the memory task also takes away questions of interest regarding the effects of word stress in NWR performance specifically for this population. Research has demonstrated that young native English speakers recognize and produce the dominant lexical meter in English which is trochaic, consisting of a strong-weak stress pattern (versus iambic weak-strong pattern (see Gerken & McGregor, 1998, for review)). Evidence from typically-developing children and adults has also shown more accurate production of stressed versus unstressed syllables, both in real- and non-words adhering to trochaic patterns (Kehoe, 1997; Morgan, Edwards & Wheeldon, 2013; Roy & Chiat).

An alternate view of these results is that lexical stress patterns can have adverse effects on NWR performance, particularly for vulnerable unstressed syllables (Gerken, 1994). Children with cognitive-linguistic SSD have been shown to have difficulties at the prosodic level by exhibiting the process of unstressed syllable deletion (Bernhardt & Stoel-Gammon, 1994; Bernthal et al., 2009; Rvachew & Brosseau-Lapre, 2012). Although weak syllable omission is considered typical in young English speakers, it is expected to resolve by age four (Bernthal et al., 2009). Children with cognitive-linguistic SSD may therefore be more sensitive than their typically-developing peers to syllable level errors in non-word repetition.

In light of the potentially conflicting effects of lexical stress on NWR performance, it is worthwhile to explore whether adding stress to the SRT significantly changes or enhances its clinical utility, as reported by Shriberg and colleagues (2009). Although Shriberg et al. argued that stimuli would be perceived as word-like even without stress cues, a design that overtly compares items with and without prosodic cues could more adeptly answer this question. Adding prosody could enhance performance of one or both groups of participants by promoting sublexical processes to facilitate accuracy. Alternatively, it could differentially affect children with cognitive-linguistic SSD by adding environments more vulnerable to syllable omission.

Theoretical bases and research hypotheses

Aside from the paucity of research regarding EFs in SSD, the justification for studying core executive functions in children with cognitive-linguistic SSD is that it significantly contributes to our understanding of what makes these children acquire

speech skills differently. This study is not primarily intended to address questions about etiology or classification of the disorder. Rather, the intent is to explore why children with cognitive-linguistic SSD appear to be frozen in their old patterns of production, while TD children transition apparently seamlessly and independently to native phonological skills.

Inhibitory control. Evidence has shown that typically-developing four-year-olds have the cognitive control to inhibit habitual or prepotent verbal and motor responses, and that this ability significantly improves during the later preschool years (Carlson, 2005; Davidson et al., 2006; Diamond & Taylor, 1996). It is possible that children with cognitive-linguistic SSD have either delayed or impaired development of inhibitory control, which results in a reduced ability to inhibit old patterns of production even when communication breaks down. Though no direct evidence has yet explored this possibility, findings from treatment studies using non-words as targets imply that these children may evince greater gains when inhibitory control of habituated productions is not required because targets are taught using new lexical forms (Cummings & Barlow, 2011; Gierut et al., 2010).

Hypothesis 1: If inhibitory control underlies the ability to inhibit early mental or physical templates when learning to produce adult-like speech targets, then children with cognitive-linguistic SSD will perform worse on tasks of inhibitory control than children with TD speech.

Cognitive flexibility. Typically-developing four-year-olds have demonstrated the cognitive flexibility necessary to alternate between different rule sets, a skill that also improves significantly with age (Davidson et al., 2006; Dauvier et al., 2012; Deak et al.,

2004; Jacques & Zelazo, 2001). Yet children with cognitive-linguistic SSD appear unable to switch flexibly from their old production patterns to the adult phonology they are exposed to. Based on suggestive experimental findings (Crosbie et al., 2009; Dodd, 2011; Dodd & McIntosh, 2008), it is worth exploring whether children with cognitive-linguistic SSD show differences in the core EF of cognitive flexibility that deter them from spontaneously adopting new phonological rules.

Hypothesis 2: If cognitive flexibility underlies the ability to spontaneously switch from early, prepotent productions to adult-like word forms, then children with cognitive-linguistic SSD will perform more poorly on tasks of cognitive flexibility than TD children.

Working memory. It has been demonstrated that four- and five-year-olds are able to effectively store and manipulate items in memory, an ability that often underlies inhibitory control and/or cognitive flexibility (Diamond, 2013; Nutley, et al., 2011). Theoretically, working memory might be involved in the process of temporarily storing word forms in the output buffer while manipulating targets during the process of correcting sounds in error. There is evidence suggesting that children with cognitive-linguistic SSD have deficits in phonological memory when compared to children with typical speech. This deficit has been proposed to relate to underlying phonological representations or differences in auditory-perceptual processes (Munson et al., 2005; Shriberg et al., 2009; Tkach et al., 2011), but it could also have implications for speech sound correction and change.

Hypothesis 3: If working memory underlies the ability to hold representations in mind while manipulating word forms in need of correction, then children with cognitive-

linguistic SSD will perform more poorly on phonological memory tasks than children with TD speech.

The relationship between EFs and speech sound accuracy. Even if no significant group differences were found in task performance, a question would remain whether executive functions relate to speech sound accuracy. Theoretically, it is reasonable to expect that children with stronger EF skills, which allow them to inhibit incorrect word forms, to transition to a mature speech system, and to hold and manipulate items in memory, would demonstrate better speech sound accuracy generally. The fourth hypothesis is as follows:

Hypothesis 4: If specific core executive functions strongly influence the transition from early mental templates to adult-like speech production, then individual differences on EF tasks will relate to speech outcome profiles. It is predicted that cognitive flexibility and inhibitory control are significantly related to speech production accuracy, while phonological memory is more related to task performance for the other two constructs.

The relationship between core EFs in preschoolers- an exploratory analysis. As summarized in the literature review, researchers are interested in how EF tasks relate to one another, not only because these findings promote the validity of new methodologies in young children, but because they further our understanding of the underlying organization of executive functions during different stages of development (Diamond, 2013; Garon et al., 2008; Miller, et al., 2012; Wiebe et al., 2008). An exploratory factor analysis using tasks from this study will examine whether the three core EFs recognized in adults are organized similarly in children, or appear to be a more unified construct during the preschool years.

Speech perception- an exploratory analysis. Researchers have proposed that speech perception might be a contributing factor in cognitive-linguistic SSD (e.g., Rvachew & Brosseau-Lapre, 2013; Rvachew et al., 1999). An experimental speech discrimination paradigm is used to assess its validity as a research tool (Rvachew, 2010). The tool was created primarily for use in therapy, but has been proposed for use in assessment as well. If the task proves valid and reliable for these purposes, the intent is to add speech perception as a control variable in other analyses.

Non-word repetition and speech sound accuracy. Shriberg and colleagues (2009) have developed a non-word repetition task in order to eliminate the confounding factor of consistent speech production errors on overall accuracy. There are several questions of interest related to this new paradigm. The first question is whether this study could replicate findings by Shriberg et al. (2009) in demonstrating performance differences between children with cognitive-linguistic SSD and TD children, when other factors, such as age or language abilities, are controlled.

Hypothesis 5: If phonological processes that are recruited in the NWR task (e.g., phonological encoding, short-term memory, etc.) are implicated in cognitive-linguistic SSD, then children in the clinical group will perform more poorly on the SRT than TD children even when other factors are controlled. In addition, it is predicted that NWR performance will relate to speech sound accuracy.

SRT and word stress. The SRT was designed as a syllable rather than non-word task, because all items are devoid of prosody. Although Shriberg et al. (2009) suggest that syllable strings are processed as non-words, it is worthwhile to experimentally examine whether the addition of word stress has a significant effect on

performance. Furthermore, it is examined whether children with cognitive-linguistic SSD would respond differently than TD children to items with and without overt stress cues.

Hypothesis 6a: If prosodic stress facilitates short-term memory for syllable strings, then all participants will have higher accuracy on stressed SRT items than on unstressed items.

Hypothesis 6b: Because children with cognitive-linguistic SSD are more vulnerable than TD children to unstressed syllable deletion, children in the clinical group will demonstrate poorer accuracy than TD children on non-stressed as compared to stressed syllables.

SRT and word length. One consistent finding in the NWR literature is that accuracy is related to word length - the longer the stimulus, the less accurate the production (Archibald & Gathercole, 2007; Munson et al., 2005; Roy & Chiat, 2004; Shriberg et al., 2009; Sutherland & Gillon, 2007). This evidence supports the use of NWR as a measure of phonological memory (Gathercole, 2006). This study asks whether there would be a length effect in the SRT and if so, whether this effect would be the same for both experimental groups.

Hypothesis 7: If the SRT is a measurement of phonological memory (among other processes), then production accuracy will decrease as length increases. It is predicted that both groups will demonstrate a similar length effect in SRT performance.

Non-word repetition, working memory and speech production. Finally, as discussed previously, researchers have argued that NWR tasks require more processes than simply phonological memory (Archibald & Gathercole, 2007; Graf Estes et al., 2007; Shriberg et al., 2009). This study enables an analysis of the relationship between NWR performance and a number of other EF tasks including working memory.

Specifically, it examines whether accuracy on the SRT would relate to performance on other EF tasks, particularly those purportedly measuring working memory. Second, it explores whether SRT performance would relate to speech sound production in the same way as measures of working memory related to speech sound accuracy.

Hypothesis 8a: If the ability to repeat non-words is, in part, a measure of phonological memory, then greater accuracy on the SRT should relate to better performance on more generalized measures of working memory.

Hypothesis 8b: Furthermore, if the SRT recruits cognitive processes in addition to phonological memory, then the relationship between SRT and speech sound accuracy will be at least partially unique from the relationship between working memory tasks and speech sound accuracy.

Chapter 2: Methods

Overview

This study was approved by the Institutional Review Board at the University of Maryland. Testing consisted of two sessions: eligibility and experimental testing. Session one included a battery of assessments intended to exclude children with low receptive language or non-verbal intelligence, borderline (i.e., low average rather than disordered) articulation skills, poor motor planning or oral-motor weakness, and hearing deficits. Session two consisted of two inhibitory control tasks, two cognitive flexibility tasks, three phonological memory tasks, a non-word repetition task, a speech perception task, and a picture-naming task to assess accuracy of speech sound production.

Participants chose a small toy after each session regardless of eligibility, while families received standardized test scores and interpretation administered by a certified speech-language pathologist. Participants' families who completed part 2 also received modest monetary compensation (\$25 for clinical families and \$10 for families of typically-developing children). Finally, as part of the consent process, parents were asked to elect whether they would allow recordings and test information to be used for teaching purposes and/or to be contributed to an international database for researchers studying child speech development.

Recruitment and screening

Participants were recruited through a variety of sources in two different metropolitan areas: Washington, D.C. and Kansas City, Missouri. Recruitment included electronically posting flyers and study information through the following sources:

- the University of Maryland Hearing and Speech department;
- referral by private and school-based SLPs to include the Park Hill school district in Kansas City (SLPs were asked to notify or send flyers home to families with children meeting the clinical description rather than pass along contact information directly to the researcher);
- listservs and websites including faith-based, daycare center, HESP alumni, and parent groups;
- word of mouth and personal referral.

Families who were interested in participating contacted the researcher by e-mail or phone. After explaining the study's intent and an overview of procedures, the researcher asked a number of questions prior to scheduling. The pre-screening questions included information about hearing infections and suspected hearing loss, percentage of English versus other languages spoken in the home, presence or suspicion of motor planning or fluency issues (to rule out apraxia of speech or stuttering), examples and parents' estimated severity of speech errors, and whether the child ever received or was currently receiving speech therapy. Prior to scheduling, the researcher also asked whether the parent would be comfortable with audio and video files being generated for research purposes. One family was ineligible to participate because the child was bilingual (using

greater than 20% in a second language) and eight families did not respond after the initial inquiry.

Part 1- Eligibility testing

The eligibility criteria were designed to establish two groups of children who differed only in articulation or phonological abilities. Prior to beginning testing, the parent signed the consent form and completed a background family and language questionnaire. Parents were invited to quietly observe the session, but more than half chose not to be present. When testing was administered at daycare centers, the researcher ensured that forms were completed before any testing was initiated.

The test battery took approximately 60-75 minutes to administer depending on the child's attention. Sequencing of tasks and exact testing locations (i.e., table vs. floor) were chosen to facilitate the child's comfort level with the researcher and to maximize the child's attention. All eligibility testing sessions were audio-recorded using a Shure SM51 microphone placed approximately eight to 12 inches from the participant and connected to a Marantz PMD600 digital recorder. The following subsections and Table 1 describe the assessments, order of administration, and inclusionary criteria.

Articulation and phonology. Group assignment was based on performance on the Sounds-in-Words subtest of the *Goldman Fristoe-2 Test of Articulation, 2nd edition* (GFTA-2; Goldman & Fristoe, 2000), a 53-item assessment that elicits 23 singletons and 16 consonant clusters in as many word positions as possible. This assessment tool is standardized for age and gender. To be eligible, children in the SSD group scored at or below the 33rd percentile on the GFTA-2 and demonstrated 0% accuracy on at least two

speech sounds across two manner classes (Morrisette & Gierut, 2002). Children in the TD group were required to score at or above the 50th percentile on the GFTA-2.

Language. All participants demonstrated no more than mild receptive language impairment defined as performance of less than 1.33 standard deviations below the mean on two standardized measures. Specifically, participants were required to score at or above the 33rd percentile on the *Peabody Picture Vocabulary Test-4th edition* (PPVT; Dunn & Dunn, 2007) to ensure they had age-appropriate vocabulary comprehension. Participants were also required to score at or above the 33rd percentile on the *Clinical Evaluation of Language Fundamentals- Preschool* (CELF-P; Wiig, Secord & Semel, 2008), Concepts and Following Directions subtest. This subtest assesses comprehension of linguistic structures for sequencing events, determining positional relationships, and describing attributes.

Expressive language testing is challenging in this population where substitutions and omissions can affect both the intelligibility of morphosyntactic markings and lexical items in general (Seeff-Gabriel, Chiat & Dodd, 2010). Two non-standardized expressive language tasks were administered: a 10-minute play session and a story-retelling task. Data from the play session (with Play-doh) were included in this study to derive measures of sentence structure and expressive vocabulary in spontaneous speech.

Non-verbal cognition. Participants were administered the Matrices subtest, a non-verbal portion of the *Kaufman Brief Intelligence Test, 2nd Edition* (KBIT-2; Kaufman & Kaufman, 2012). This subtest, in which the child matches one concrete picture or abstract shape to a target among a field of five like items, assesses children's abilities to

find relationships between items and to complete analogies. All participants were required to score at or above the 33rd percentile for their age.

Hearing. Participants were also required to pass an audiometric screening at 25 db HL at 500, 1000, 2000, 4000 and 8000 Hz (American National Standards Institute, 1991) in at least one ear.

Oral-motor skills. Participants were required to pass an oral-periphery mechanism screening using a standard protocol developed by Robbins and Klee (1987), which is intended to rule out speech errors due to structural abnormalities and/or apraxia of speech. The screening assesses oral-motor range of motion, stimulation of four phonemes in isolation that vary by place of articulation and manner, sequential and diadochokinetic production, and repetition of a list of 14 one- to four-syllable words. Scoring was based on a pass/fail criterion; any child who demonstrated structural abnormalities or planning difficulties was excluded.

Table 1. Eligibility testing protocol and criteria

Order of testing	Assessment tool	Eligibility Criteria
1, 2 or 3	CELF-P	> 33 rd percentile
1, 2 or 3	KBIT	> 33 rd percentile
1, 2 or 3	GFTA	SSD: ≤ 33 rd percentile TD: ≥ 50 th percentile
4	PPVT	> 33 rd percentile
5	O-M screening protocol	Non-standardized- clinically appropriate anatomy and physiology
6	Frog story Play session	N/A (experimental expressive language variables)
7	Hearing screen	25 db HL at 500, 1000, 2000, 4000 and 8000 Hz unilaterally
N/A	Parental report of monolingualism	>80% English-speaking

Participants

Part 1- eligibility cohort. A total of 82 four- and five-year-old children (36 female) were tested for participation in this study. Figure 1 provides a summary of participants who did and did not meet eligibility requirements. Seventeen of those children (7 female) did not meet the eligibility criteria for the following reasons:

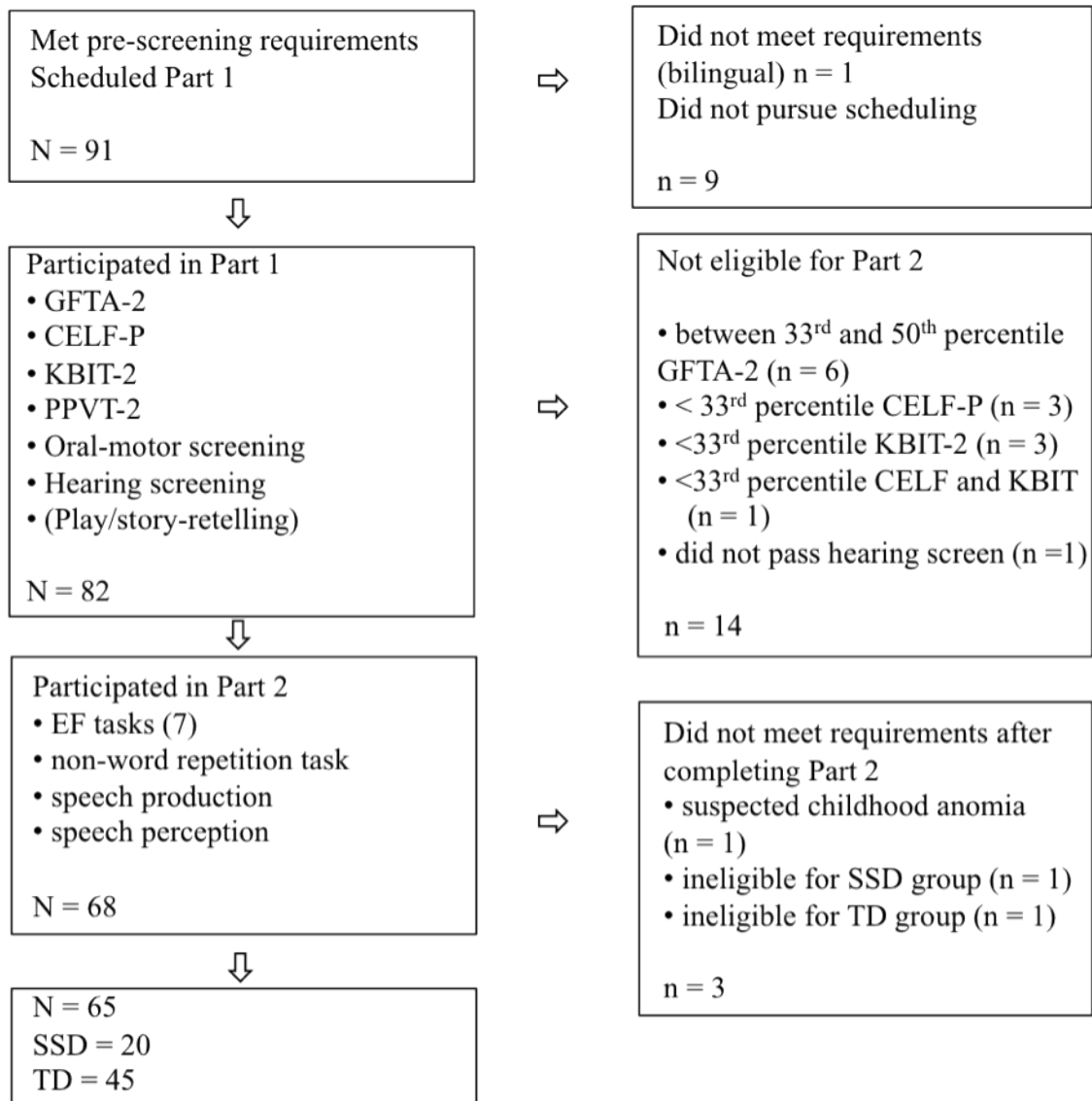
- three children did not exceed the cut off score on the CELF-P subtest for receptive language;
- three children did not meet the criterion on the KBIT-2 test of non-verbal intelligence;
- one child failed to meet the criteria on both the CELF-P and KBIT-2;
- six children scored in the low typical range, between the 33rd and 50th percentiles, on the GFTA-2;
- one child failed the hearing screening due to bilateral deficits.

In addition, three children completed experimental testing, but were later excluded for the following reasons:

- one child from the typical speech group presented with suspected childhood anomia as evidenced by significant difficulty in naming experimental stimuli that affected task performance;
- one child was excluded from the clinical group when within-category distortions were classified as accurate according to the Percentage Consonants Correct- Revised (PCC-R) measure (Shriberg, Austin, Lewis, McSweeny & Wilson, 1997; see coding section for further discussion regarding this distinction);

- one child who was originally included in the typical speech group was later excluded on the basis of below average or borderline SSD, as indicated by his performance on the Picture Naming Task (PNT; Preston & Edwards, 2010) in Part 2.

Figure 1. Flow chart of testing with numbers of participants



Part 2- experimental cohort. Sixty-five (65) four- and five-year-old children met the criteria for and completed experimental testing. Twenty children (9 female) were in the clinical group and 45 (20 female) were in the typically-developing speech group. As intended, the groups were well matched on all standardized assessments except for articulation and phonology. As can be seen in the test score data (listed in Table 2), on average, participants in both groups scored above the mean on receptive language and non-verbal cognitive measures.

Table 2. Standardized test scores

	SSD (n = 20)			TD (n = 45)		
	Mean	SD	Range	Mean	SD	Range
GFTA-2 Standard Score	86	9.3	64 – 99	114	3.1	107 - 121
Percentile Rank	17	8.8	3 - 33	80	10.0	54 - 96
CELF-P Standard Score (Concepts and following directions subtest)	110	7.7	95 – 125	110	11.9	95 – 145
Percentile Rank	71	16.0	37 - 95	71	20.3	37 - 99
PPVT-4 Standard Score	115	9.9	100 – 136	123	13.6	98 – 152
Percentile Rank	79	14.2	50 - 95	86	15.3	45 - 100
KBIT-2 Standard Score (Matrices subtest)	108	8.2	94 - 120	111	11.5	94 – 138
Percentile Rank	68	18.2	34 - 91	71	20.6	34 - 99

Demographic characteristics. Family history and other demographic variables were collected from a questionnaire that caregivers completed during the initial session,

which did not include race and ethnicity information. Questionnaires were optional and some parents chose not to complete particular questions. The data (listed in Table 3) included a mix of continuous, categorical and binary variables to include the following:

- ages of acquisition of key developmental milestones reported in months;
- parental level of education reported in years;
- family history of speech, language, and cognitive-communication problems scored as a binary measure. Positive history included speech disorder or problems with intelligibility, delayed language, mental retardation, autism spectrum, and dyslexia, but excluding ADHD for both first- and second-degree relatives;
- history of surgical or medical information scored as a binary measure. Positive history included Pressure Equalization (P-E) tubes and surgeries affecting respiratory or oral structures, but not including tonsillectomy;
- history of feeding problems used as a binary measure. Positive history included parental report of difficulty with bottle-feeding or latching, poor weight gain, and reflux;
- ear infections treated as a binary variable (i.e., positive versus negative history) rather than numbers of infections.

Table 3. Demographic variables by group

	SSD (11 male; 9 female)			TD (25 male; 20 female)		
	Mean	SD	Range	Mean	SD	Range
Age in months	60.5	6.5	48 – 71	59.8	6.3	48 – 71
Age in months of first word	13.0	4.2	5 – 24	13.1	4.2	8 – 24
Age in months of short phrases	20.8	5.6	11 – 30	19.2	5.2	9 – 30
Age in months of walking	13.1	2.4	10 – 19	12.3	1.7	8 – 17
Highest level of maternal education in years	16.0	2.1	12 - 20	17.1	1.4	12 - 20
Highest level of paternal education in years	16.1	2.2	12 – 20	16.5	4.1	12 - 20
Proportion with positive family history of speech/language problems	.8	N/A	N/A	.3	N/A	N/A
Proportion with surgeries/conditions involving speech and hearing	.1	N/A	N/A	.1	N/A	N/A
Proportion with history of feeding problems	.2	N/A	N/A	.1	N/A	N/A
Proportion with ear infections	.2	N/A	N/A	.17	N/A	N/A

Values are reported as group means. Standard deviations are in parentheses. N/A reported for standard deviation and range for binary variables. Education levels: 12=high school; 14=some college; 16=bachelor's; 18=master's; 20=post graduate.

Parental dialect was consistently reported as either mid-Atlantic or Midwestern. Only one parent out of 65 responded that she was concerned about her child's communication development. Finally, fewer than 10% of children were exposed to a second language. Two children in the TD group were exposed to less than 10% (of Hindi

and Spanish), four children in the SSD group heard less than 5% of a second language (French, Spanish, German, Italian, and Portuguese), and one child in the SSD group was exposed to approximately 18% Italian.

Part 2- Experimental testing

Overview. Experimental sessions took place approximately two weeks after eligibility testing (mean = 12.3 days; range = 1-34 days). Sessions were audio recorded using the experimental set up as described in Part 1. One experimental task was also video recorded using a Sony 120x digital Handycam. For video recording, equipment was aimed diagonally from behind the participant in order to view the laptop screen while recording verbal responses. Testing sessions lasted between 90 to 120 minutes depending on the child's level of attention.

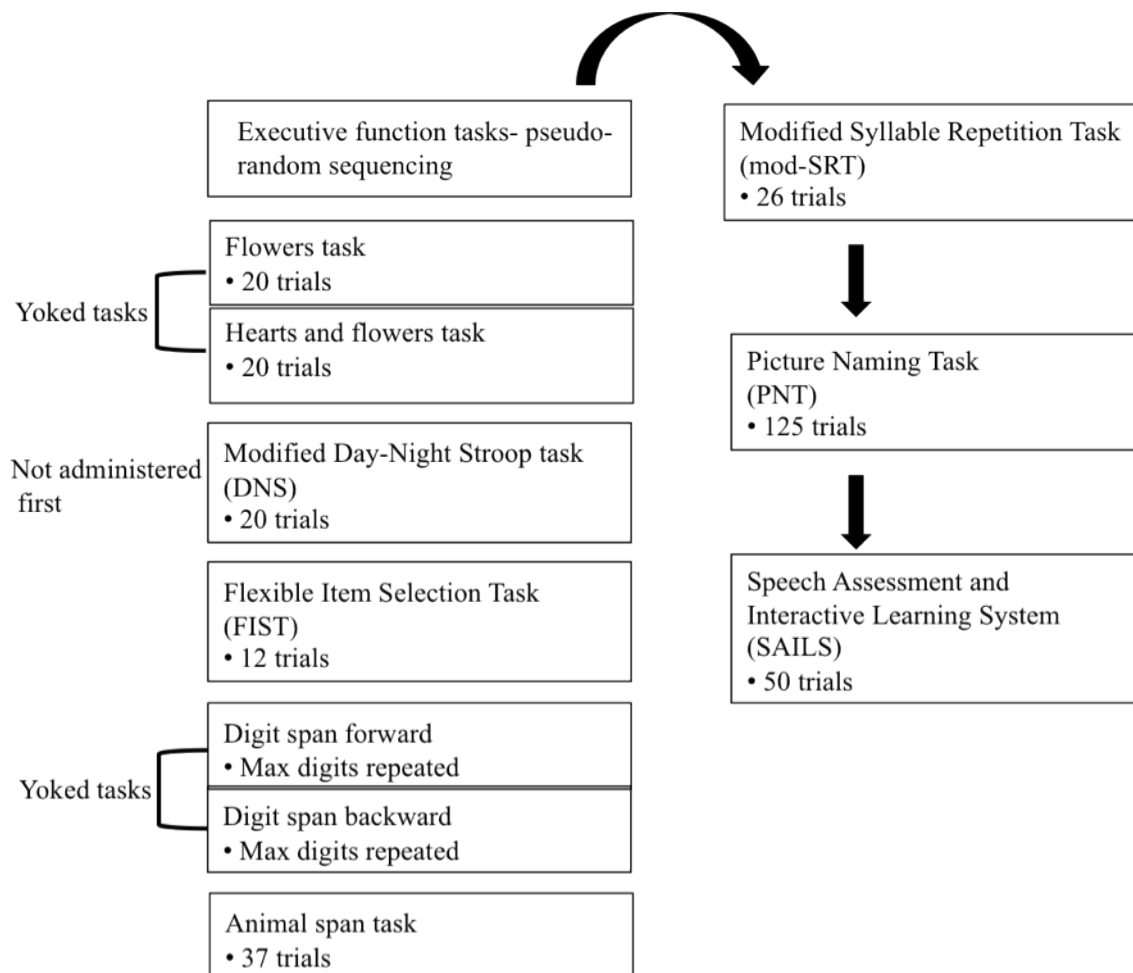
The order of administration of tasks used a set sequence. EF tasks were pseudo-randomized with the following exceptions:

- the modified Day-Night Stroop task (DNS; Pasalich, Livesey & Livesey, 2010) was most successfully administered later in the testing, once children were used to attending to the “games;”
- the forward and backward digit span tasks and the Flowers, and Hearts and flowers tasks were yoked because of step-wise sequencing (i.e., in order to complete the second task, children had to perform the first).

After completing the EF tasks, participants were administered the non-word repetition, speech production, and speech perception tasks in that order. The task sequence was chosen so that the most cognitively demanding tasks were administered earlier, before

fatigue became a potential factor. Figure 2 summarizes the tasks by sequence order and number of trials per task.

Figure 2. Order of experimental tasks and number of test trials



Modified Day-Night Stroop task (DNS; Pasalich et al., 2010). This task was intended to assess inhibitory control, specifically verbal response inhibition. In this task, children are asked to verbalize semantically opposite labels to the images they see. Though it is similar to other pre-literate Stroop tasks such as the Red dog/Blue dog task (Gerstadt, Hong & Diamond, 1994), Pasalich and colleagues have shown this task to be

more challenging for 4- to 6-year-old children thus avoiding ceiling effects. The task in the current study was adapted from Pasalich et al. (2010), with two main differences: 1) four more test trials were added to the original 16, and 2) children were given 300 msec longer to respond, for a total of 2300 msec. The greater number of trials was intended to further tax executive control, which has been shown to wane over successive trials (Diamond & Taylor, 1996). The slightly longer response time was judged to be a more appropriate time interval especially for the younger participants, during piloting of task design.

Stimuli consisted of four Microsoft Clip Art 3 x 1.5-inch characters, which included a girl, a boy, a dog and a cat. Introduction and practice blocks as well as two orders of 20 trials were created. The experimental orders were pseudo-randomized such that each character occurred once in every four trials, but no character was presented twice in a row. The two orders were counterbalanced across participants. The task was administered on a MacBook Pro using the program Psyscope (Cohen, MacWhinney, Flatt & Provost, 1993). Introduction and practice trials were untimed, but experimental trials automatically advanced after a 2,000 msec display followed by 300 msec fixation cross.

Participants were introduced to the task with the following directions: *“I want to introduce you to my friends. They have very silly names.”* Each character was displayed in turn as the researcher introduced the name with its opposite label (e.g., [picture of dog] “this is Cat;” see Figure 3 for illustration). After the introduction, participants were required to demonstrate comprehension of the task by verbalizing opposite labels during a practice set, which consisted of each of the four characters. If participants did not

achieve 100% accuracy during the practice trials, the researcher repeated the introduction and practice trials with the child a second time.

During the experimental test, responses were scored on-line as well as video-recorded and scored a second time to ensure intra-rater reliability. If children were struggling with the task, the cue “*remember their **names***” was offered no more than two times. General encouragement such as “*you’re doing a great job*” and “*almost done*” was provided for at least every third trial. Children’s responses were counted if voice onset was initiated prior to the next trial being displayed. Responses could overlap with the next trial’s fixation cross (2300 msec maximum), but not with the successive trial image.

Figure 3. DNS task example.

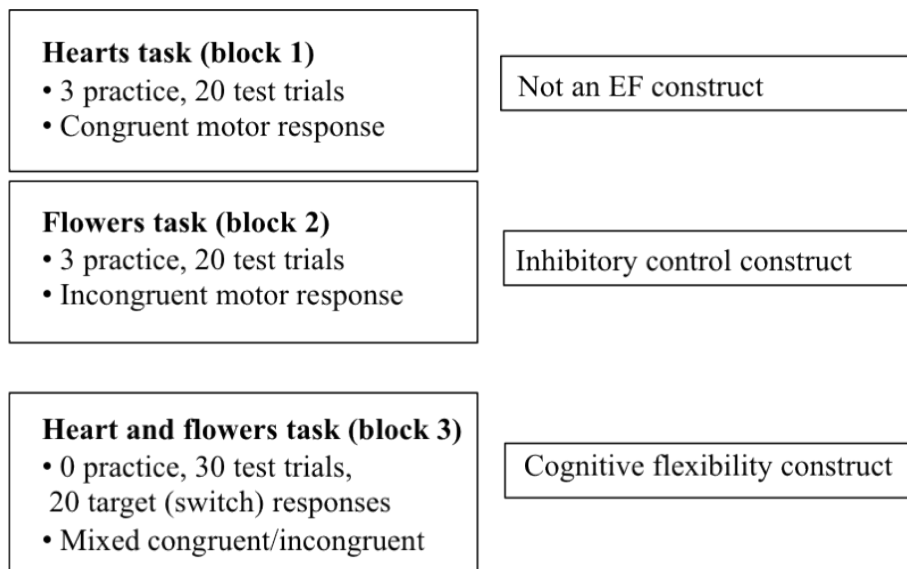


Hearts and Flowers task. Hearts and Flowers (Diamond, 2013) is the child-friendly version of the Dots task introduced in Davidson et al. (2006). In this task children are asked to press a key on either the same or opposite side of the keyboard as the symbol they see on the screen. There are 3 blocks in this task that are administered in a set order based on the following design (see Figure 4 for a summary of the design).

1. In the first block (the Hearts task), the prepotent motor response is reinforced when the child is asked to press a key on the same side as the stimulus (a heart)

- being presented. This task does not target an EF construct *per se*, but instead is used as a precursor for the Flowers task.
2. The second block (the Flowers task) tests the strength of the Simon effect; the child is asked to inhibit the prepotent and now habituated congruent response and instead press a key on the opposite side to the stimulus (a flower). The second block was used to assess inhibitory control.
 3. Finally, the third block (the Hearts and flowers task) mixes the two types of responses, thus testing the child's flexibility in switching motor responses to match the stimulus (heart or flower) being presented. This block was used to test cognitive flexibility and will be discussed in the next section.

Figure 4. Hearts and Flowers task design



Flowers task- blocks 1 and 2. Stimuli consisted of one blue heart and one blue flower from Microsoft shapes that were 3/4-inch in diameter and displayed on a white

background. The [a] and ['] keys on the researcher's laptop were each covered by a different sticker, which served to mark the response keys. Although reaction times are known to be less accurate with key presses than with button box entries (by approximately 9-12 msec; L. Filipen, personal communication, January, 27th, 2013), because this study compared between-group differences, key presses sufficed.

The first two blocks (or tasks) each consisted of 3 practice trials and 20 test trials. For these tasks, the stimulus occurred 10 times each on the right and on the left side of a fixation cross positioned 2.5-inches from center. Trials within each task were pseudo-randomized in a set order such that no more than two consecutive trials on one side were presented. Two orders were created that were counterbalanced across participants.

The laptop was set up so that the response keys were within easy reach of participants' left and right index fingers; participants were shown how to rest their arms on the laptop to stabilize their hands. For the Hearts task, the following instructions were given: *"When you see a heart, push the button on the same side as that heart. So you if you see a heart on this [point to the left] side, you push this [the left] button. If you see a heart on this [point to the right] side, you push this [the right] button. Let's practice."* Participants had to respond to at least two out of the three practice trials correctly. If they did not, a second round of instructions and practice trials was administered.

After successful completion of the practice trials, participants were told *"Now the game goes kind of fast. Make sure you see the heart before you decide which button to push."* For the 20 test trials, participants were given 1500 msec to respond with 500 msec inter-stimulus interval (ISI) marked by a fixation cross. Participants were offered feedback every three to five trials that was intended to promote attention (*"keep going,*

you're almost done”). Though accuracy and reaction times (RT) were collected from the Hearts trials, the data were not analyzed for purposes of this study.

For the Flowers task, participants were given the following instructions: *“Now this time the rules of the game are different. When you see a flower, push the button on the **opposite** or **other** side as that flower. So you if you see a flower on this [point to the left] side, you push this [right] button. If you see a flower on this [point to the right] side, you push this [left] button. Let’s practice.”* Again, after correctly completing two of the three practice trials- with a second administration to meet the criterion for accuracy if necessary- participants were told *“this game goes pretty fast. Wait till you see the flower before you press the key. Here we go.”* Participants then completed 20 test trials. General encouragement was provided every three to five trials, in addition to no more than two cues of *“remember which side”* for the duration of the task. The software program Psyscope recorded response accuracy and reaction times.

Hearts and flowers task- Block 3. The final task combined hearts and flowers stimuli to assess the construct of cognitive flexibility. This task consisted of 30 trials, only 20 of which were switch trials. Switch trials required the participant to change rule sets from congruent to incongruent or incongruent to congruent responses. Trials that did not require a switch in the response pattern were not calculated for purposes of this study (Deak et al., 2004). Similar to the first two tasks, orders were pseudo-randomly created so that the same shape occurred no more than twice consecutively. Left- and right-side responses were counter-balanced such that switch trials occurred an even number of times on each side. Two orders were created that were counterbalanced across participants.

Participants were given the following instructions: *“Now I’ve got a real challenge for you. Remember when you saw a heart in the first game and you pressed the button on the same side? Then remember when you saw a flower in the second game and you pressed the button on the opposite side? Now we’re going to mix up the hearts and flowers. It goes kind of fast. Don’t worry if you get some wrong- just keep going. Are you ready?”* As per Davidson et al. (2006) there were no practice trials for this block.

Participants completed the 30 trials with general encouragement and no more than two cues of *“remember which side!”* Participants had 2,000 msec to respond followed by a 500-msec ISI indicated by a fixation cross. Again, accuracy and reaction times were generated in Psyscope.

Flexible Item Selection Task (FIST). The second measure chosen to assess cognitive flexibility was the FIST, first published by Jacques and Zelazo (2001). Dodd and colleagues (Dodd & McIntosh; 2008; Crosbie et al. 2009; Dodd, 2011) used this task to test children with SSD. The current task used 48 4x6-inch cards with pictures that varied by shape (*fish, car, sock*), size (*small, medium, large*), number (*one, two, three*) and color (*yellow, blue, red*). For example, one card consisted of two small red cars, while another card consisted of three medium blue fish. For each of the 12 test trials, there was a designated set of three cards. In addition to the 48 test trial cards, there were three sets of four cards designated for one demonstration and two practice trials. Unlike test trial cards, demonstration and practice card sets consisted of two sets of matching cards (e.g., two cards with one small red fish and two cards with three medium blue cars).

To introduce the task, participants were shown the set of four demonstration cards with the following instructions: *“I have four cards here. I’m going to point to two cards*

that go together because they're both _____ [size, shape, color and number]. You see?
*Now I'm going to pick out two **other** cards that go together because they're both _____*
[list of all four attributes].” After the demonstration, the researcher laid out the first practice set and asked the participant to point to first one and then another pair of cards that ‘go together.’ After each selection, the researcher highlighted the four attributes that each pair of items had in common. Finally, participants were administered the second set of practice items for which they were cued to pick one and then a second selection. The researcher did not review the attributes on the second practice trial. Participants were required to score 100% accuracy for two selections (i.e., pair one and pair two) on each of the practice trials.

After the demonstration and two practice trials, test trials were administered (see Figure 5 for an example). Half of the participants were administered the task in the order described by Jacques and Zelazo (2001), while the other half of participants were administered the order in reverse. The following instructions were given prior to starting the test trials: *“Now, instead of four cards, I’m going to show you just three cards. I want you to do the same thing as before. Find two cards that go together.”* After making their first selection, regardless of accuracy, participants were prompted to *“point to two **other** cards out of these three that go together.”* Participants were given general feedback at least every other trial (*“you’re doing a great job”*). Participants were also given no more than two specific cues (*“think about all the ways that things go together like we talked about”*). Participants who pointed to only one item for the second selection were also asked *“does that card go with anything else?”* All responses were written down on-line but scored off-line.

Figure 5. FIST example



This task is designed to test two separate, but related executive function constructs: rule abstraction and cognitive flexibility. The ability to identify a similar pair of items based on one salient dimension is thought to fall under rule abstraction, while the ability to select a second pair based on a different dimension requires both rule abstraction and cognitive flexibility (Crosbie & Dodd, 2009). Though accuracy of both participants' first and second selections was noted, the relevant measure in this study was total number of accurate second selections in which the first choice was accurate. That is, if the first selection was correct, the second selection was scored as accurate or inaccurate. If the first selection was incorrect, the second selection was automatically counted as incorrect.

Digit span task overview. As summarized in Figure 2, digit span forward and backward tasks were always administered consecutively. The digit span forward task assesses verbal short-term memory capacity, but the backward span task is proposed to tap into both capacity and mental manipulation processes (Diamond, 2013). Both span tasks are commonly administered subtests from the Wechsler Adult Intelligence Scale-III (WAIS-III, Wechsler, 1991).

Stimuli were presented as audio files rather than by live voice in order to ensure consistency of rate, prosody, and volume. A female graduate student from the Language Development Lab at the University of Maryland who spoke with a mid-Atlantic dialect recorded the stimuli. The speaker produced three repetitions of digits *one* through *ten*, excluding *seven* because it is not monosyllabic (Adams & Gathercole, 1996) into a Shure SM51 microphone in a sound-attenuated booth. Files were digitized via a 16-bit analog-to-digital converter at a 44.1 kHz sampling rate. One production per digit was selected based on naturalness of rate and prosody and then all stimuli were adjusted to a consistent amplitude.

Two orders of randomly sequenced digit strings and three practice trials were created for the tasks. No digit was repeated in a single string. Each order included two random sequences of digits per level (e.g., two sequences of two digits, two sequences of three digits, etc.) with a maximum level of six digits. All strings were created with two seconds of silence between numbers. Orders were counterbalanced across participants such that half the children were administered order one for the forward condition and order two for the backward condition, while the other half of participants were given forward and backward conditions with orders two, then one.

Forward digit span. Participants were given the following instructions: “*You are going to hear some numbers and I want you to repeat them exactly as you hear them.*” After successfully repeating one practice trial consisting of a single digit, participants were administered a second practice trial of two digits. Though children were instructed to wait for the end of the sequence prior to repeating the stimuli, waiting proved to be difficult particularly for younger participants. Some children waited for the

researcher to lower her hand from a quiet signal position, while other participants were taught to use their own hands to cover their mouths until the researcher signaled them to repeat.

After the practice trials, regardless of accuracy, participants were administered test trials starting with two-digit strings. If the child correctly repeated the first string, the researcher moved to the next level. If the first string was incorrect, the child was presented the second string at the same level. This procedure continued until the child was unable to repeat either string at a particular level. Participants were given only general encouragement during the task. Accuracy was defined as the maximum level at which a participant successfully repeated one digit string out of two possible trials.

Backward digit span. Participants heard instructions for this task as follows: *“Just like last time, you are going to hear some numbers. However, this time I want you to repeat them back to me in the **reverse** order or **backwards** from how you hear them.”* Participants were given one demonstration trial followed by one practice trial with two digits each. During the demonstration trial, the researcher further illustrated the task by using directional arm movements to conceptualize forward versus backward sequencing. Corrective feedback was provided on practice trials. No accuracy criterion was required before advancing to test trials. Cueing and accuracy were the same as described in the forward span task; both scores were annotated on-line.

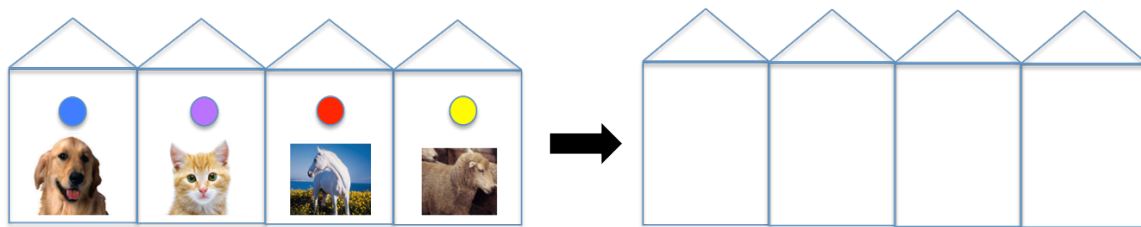
Animal span task. The third phonological working memory task was modeled after Willoughby et al.’s (2012) working memory span task. This task includes an articulatory interference component, which is intended to further tax working memory (Fatzer & Roebbers, 2012). This task differed from the original study in two aspects: 1) the

number of stimuli was increased from 18 to 37, and 2) the maximum number of items recalled in a trial was increased from four to five. The stimuli included six Microsoft Clip Art animals (*cat, dog, cow, sheep, pig, and horse*), six colored dots (*red, blue, purple, green, yellow, and orange*), and a line drawing of a house. Two orders were created in Power Point, each of which included a familiarization slide with all animals and colors, one practice trial, and 12 test trial slides. The two orders were counterbalanced across participants. On practice and test slides, each house contained exactly one animal and one colored dot. The test trials included:

- two trials each of one- and two-houses;
- three trials each of three- and four-houses;
- two trials of five-houses.

On alternating slides, the same number of houses was pictured as in the preceding slide, but without any animals or colors (see Figure 6 for an illustration).

Figure 6 Animal span example



This task began by having participants provide all 12 animal and color names to ensure 100 percent familiarity with test items. Children were allowed to use alternate names that were more natural to them (e.g., kitty/cat, lamb/sheep). The instructions for the practice task were as follows: *“I want you tell me first the color and then the animal you see in this house.”* After naming the color and animal, the slide was advanced to an empty house and participants were prompted to *“name what animal used to be in the*

house.” Throughout all 12 test trials, participants were prompted to name first the color then the animal in each house, but only to recall the animal names. Order of recall did not affect accuracy. A puzzle activity or game was interspersed between trials to facilitate attention during the task. The total number of animal names successfully recalled was recorded on-line.

Speech sound production task. The Picture Naming Task (PNT; Preston & Edwards, 2010) is a 125-single-word naming task that features each speech sound at least twice in every word position (See Appendix A for list of stimuli). Two advantages of the PNT over other tools cited in the literature are 1) it includes a number of multi-syllabic words and consonant clusters for a more complete picture of the child’s sound system, and 2) it is of average length, thus not consuming too much time, yet providing sufficient data for analysis. The PNT yields a total of 480 consonants if all word forms are produced as intended.

Participants were asked to name the picture or complete a sentence with the target word pictured on the laptop screen. In the event the child was unable to spontaneously name the item, the target was elicited via delayed imitation (*e.g.*, “*This is a parachute. What is it called?*”). Studies have shown that delayed imitation is a good reflection of spontaneous articulation performance (Paynter & Bumpas, 1977; Templin, 1947; but see Johnson & Somers, 1978). Variations in word form such as affixes were not corrected. For half the participants, PNT items were presented in reverse order.

A painting activity was interspersed throughout the task as an incentive to continue naming pictures; every 15 to 20 items, children chose a paint color to add to a

spinning picture. Two children required incentives beyond the painting activity (one responded to candy, and the other played an additional game). Other than general cueing as needed, no feedback about accuracy was provided on test items. Accuracy was annotated during the session particularly for those phonemes that were difficult to distinguish in audio-recordings (e.g., /f/ versus /θ/), but coding was otherwise completed off-line. On average, the PNT took 20 minutes to administer.

Non-word repetition task overview. As discussed in the introduction, classic non-word repetition paradigms are often modified for children with SSD; without some sort of modification, phonological memory performance is confounded with substitution or omission errors that are part of the child's speech sound system rather than a phonological processing deficit. Shriberg and colleagues (2009) created the Syllable Repetition Task (SRT) to avoid the necessity of adjusting scores *post hoc* in this clinical population. Although preliminary results from the SRT were promising, Shriberg et al. (2009) identified a number of concerns that could be modified in the task design. First, participants with typical language who are closer to age six may demonstrate ceiling effects. Second, interference among items due to the limited number of phonemes used in the stimuli might negatively affect accuracy for reasons other than memory. Finally, the SRT is relatively free of stress markings that indicate word-likeness and is therefore less likely to capture effects from syllable structure errors.

To address these concerns, a modified version of the task was created that consisted of two conditions: stressed and equal stress. The conditions were pseudo-randomly ordered rather than blocked. This design not only enabled analysis of syllable

structure accuracy, but also may have alleviated some interference effects by adding variation across items. The task also included items of increasing length (from two- to five-syllables) for two reasons: 1) to address questions related to two-syllable accuracy, and 2) to alleviate ceiling effects.

Modified-SRT stimuli. The task consisted of 13 items for each condition, which included four 2-syllable, three 3-syllable, four 4-syllable, and two 5-syllable items (see Appendix B for the list of stimuli). In the stressed condition, six items were created with a trochaic stress pattern and seven with an iambic pattern, which were evenly divided across syllable levels (with the exception of the 3-syllable level that had two iambic items). Stimuli were recorded using the same procedures described under digit span tasks, only by a different female lab member with mid-Atlantic dialect. All items were produced with falling intonation and were recorded a minimum of three times in order to select the best tokens for the test trials.

The speaker recorded each of the equal-stressed stimuli immediately after listening to the same item from the Phonology Project SRT that is available online (Shriberg & Lohmeier, 2008). After recording repetitions of each SRT stimulus, the speaker produced repetitions of a stressed condition item with the same number of syllables (ex. recording /'ma'da/ followed by /nə'ba/). The student was instructed to match the rate of the original SRT item as closely as possible by tapping one hand to the syllables, which ensured that the rate of syllable onset was constant across condition. Paired stimuli across conditions were not matched in length; though similar, equal-stress items were consistently longer than stressed items. Across conditions, 2-syllable items had an average duration of .799 msec (range .66 - .867 msec), 3-syllable an average of

1.218 msec (range 1.02 - 1.372 msec), 4-syllable an average of 1.630 msec (1.508 – 1.747 msec), and 5-syllable items averaged 1.860 msec (range 1.764 – 1.931 msec).

Stressed stimuli were marked by both intensity and vowel quality, two cues which have been found to be relevant in interpreting lexical stress (Morgan et al., 2013; Zhang & Francis, 2010). Specifically, vowels in stressed syllables had an average intensity of 79 dB as measured in PRAAT (range 70.4 – 80; Boersma, & Weenink, 2009), while vowels in unstressed syllables averaged 75.6 dB (range 62.2 - 80). Morgan et al. (2013) found intensity and not duration to be a critical acoustic cue for lexical stress. Stressed syllables in this condition consisted of /a/ (which was also the only vowel used in the original SRT stimuli) while unstressed syllables were marked by schwa. Appendix C provides an example in PRAAT (Boersma, & Weenink, 2009) of paired stimuli across the two conditions that are similar in rate of syllable onset, but distinguished by intensity (vowel quality is not visualized).

Modified SRT procedure. Participants were given the following instructions: *“I’m going to play you some silly words and I want you to tell me **exactly** what you hear. So, for example, if you hear the word “teepa,” you say _____. If you hear the word “peeku,” you say _____.”* Practice items were administered a second time with corrective feedback if participants did not accurately repeat them. Test trials were presented binaurally through Sennheiser HD-201 headphones at a comfortable listening level. Half of the participants received one of two pseudo-randomized orders, in which stimuli within a syllable level were randomized but no more than two items from one condition were presented consecutively. Children were given general encouragement throughout

the task and no more than four specific cues stating “*remember to repeat **exactly** what you hear.*”

Two-thirds of the participants completed a puzzle along with the task, which was intended to serve as a reinforcing activity and to promote attention; however, the puzzle was initiated only after the first 25 children had already completed testing. For the latter group of participants, children received one puzzle piece after repeating two stimuli. This difference in how the protocol was administered will be discussed further in the results and discussion sections.

Speech Perception. The software program Speech Assessment and Interactive Learning System (SAILS; Rvachew, 2010) was used to provide a measure of global speech sound discrimination skills. This program provides a two-alternative forced choice (yes/no) paradigm in which the child hears a word through headphones and points to the picture that matches the word she hears (e.g., ‘soap’ vs. “X”). The program includes practice trials featuring maximally distinctive contrasts such as ‘soap’ and ‘moap’ to teach the discrimination paradigm, and assessment trials for eight phonemes in word-initial position.

The test trials for each of the eight phonemes consist of 10 items each, half of which are different tokens. Test trials are presented in random order and produced by a variety of speakers. The software also includes varying levels of difficulty depending on the target phoneme. For practice and test trials, the child’s response is entered into the program, which not only adds an on-screen puzzle piece reinforcer for the child, but also discretely shows the accuracy of the child’s response to the administrator. The program

later generates a spreadsheet of trial-by-trial responses as well as an overall accuracy score.

For the purposes of this project, four speech sounds from the last stage of phonemic acquisition (/l/, /r/, /θ/ and /s/; Shriberg, 1993) as well as one mid-stage affricate /ʃ/ were assessed. Although SAILS is intended to probe and treat speech discrimination of specific sounds produced in error, it was used in this study as a more general measure of speech sound discrimination. This method of sampling across sounds to generate an overall measure of speech perception is based on a similar method used to measure general consistency of errors (Tyler et al., 2003).

Participants were taught the discrimination paradigm using the following instructions: *“You’re going to hear different people say a word. Some of the people speak well and some people don’t speak well meaning they don’t say their sounds well. Your job is to decide if the word sounds good/right or bad/wrong.”* The choice of terminology ‘good’ and ‘bad’ versus ‘right’ and ‘wrong’ was child dependent (i.e., whichever terms facilitated the child’s comprehension of the task). After the instructions, participants completed a minimum of four practice trials of initial /k/ with 75% accuracy; more practice trials were administered with corrective feedback until participants achieved the criterion for accuracy. Clinical participants who produced initial-/k/ in error (three total) were administered practice trials with initial /f/. Participants wore Sennheiser HD-201 headphones for all practice and test trials.

Participants completed five sets of 10 discrimination trials presented in random order. All phonemes were administered on difficulty level two with the exception of /l/,

which does not offer a difficulty level higher than level one. Only general feedback was given throughout the task.

Coding and Analysis

Executive functions. Participants' responses were recorded on-line by either the researcher or by Psyscope, but actual scoring was calculated off-line. For the digit span tasks, the last accurately repeated digit string functioned as the raw score. For the animal span task, every successfully recalled animal in a trial regardless of the order in which it was named counted toward the total raw score. For the FIST, the number of correct second choice responses for which the first choice was also correct functioned as the accuracy score. The modified DNS task was re-scored by the researcher using video-recordings to ensure that delayed responses (i.e., responses initiated just before the next trial onset) were accurately coded; for four participants, the video-recordings were inaccessible (n = 3 operator error; n = 1 recorder malfunction). The total number of correctly named responses functioned as the raw score. Finally, accuracy and reaction times of the Flowers task and of the Hearts and flowers task were calculated from Psyscope output. Button presses under 200 msec (as per Davidson et al., 2006) were removed from both accuracy and RT measures, which resulted in the exclusion of 3.7% and 1.7% of the data from the Flowers and Hearts and flowers tasks respectively. Inaccurate responses were not included in RT calculations.

Because maximum accuracy scores ranged from 12 to a possible 37 depending on the task, raw scores were converted to z-scores for particular analyses. Z-scores are a useful means of normalizing tasks for comparison by using population data to standardize

values. For purposes of this study, raw scores from the TD group were used to calculate population mean and standard deviation for each task.

Spontaneous language production. A minimum of 40 child utterances from each play session were transcribed and analyzed using the software program CLAN (MacWhinney, 2000). Single word utterances were not included in the total. Language samples were transcribed from the beginning of each session. The command DSS (Developmental Sentence Score) was run to obtain a measure of syntactic complexity, where higher values indicate greater competence with linguistic structures. The command VOC-D was used to measure vocabulary diversity; this algorithm is similar to type-token ratio, but is proposed to control for variations in sample size (see MacWhinney, 2000, for review).

Speech production. Responses from the PNT were linked and broadly transcribed using the freeware software Phon (Rose et al., 2006). Phon is a transcription and analysis program with several benefits including:

- easily accessible International Phonetic Alphabet (IPA) symbols;
- data entry for the orthographic, target phonetic, and actual phonetic productions;
- the use of templates for tasks that have a set number and sequence of target items;
- a number of analyses such as Percent Consonants Correct (PCC) and searches of deletions/substitutions/epentheses by syllable position.

One template for each order of the modified-PNT (forward and reverse) was entered into Phon and used to transcribe each participant's productions.

Although diacritic marks and allophonic variations were often transcribed or noted, they did not factor into the primary measure of interest, Percent Consonants Correct-Revised (PCC-R; Shriberg et al., 1997). Distortions that were judged within the phonemic boundary for the target phoneme were counted as accurate. Specifically, variants of the alveolar fricatives /s/ or /z/ were counted as distortions and thus accurate, unless judged to be other English dental or palatal fricatives; PRAAT (Boersma, & Weenink, 2009) was used to visualize the noise distribution of the signal. Liquids /r/ and /l/ were played in isolation using PRAAT in order to judge whether they fell within or across phonemic boundaries. For example, items that sounded like a distorted /r/ were counted as accurate, whereas those that sounded like a /w/ were not.

There were several rationales for using the PCC-R measure, which counts distortions as accurate productions rather than errors. As pointed out by Preston and Edwards (2010), evidence suggests that distortions may be motorically-based rather than cognitive-linguistic error patterns. Although correcting both motoric and cognitive-linguistic errors may require core executive functions, the purpose of this study is to examine children with cognitive-linguistic SSD rather than a subtype of SSD remarkable for residual or persistent errors on sibilant (/s/) and/or liquid productions (Shriberg et al., 2010). In addition, distortion errors constitute a relatively small proportion of errors in the SSD literature (Gruber, 1999; Preston & Edwards, 2010). Finally, Gruber (1999) proposed that distortions may represent the transition from substitution or omission to accurate production, thus indicating that participants are in the midst of correcting their productions.

In addition to the PCC-R criteria, a number of additional coding conventions included the following:

- items with greater than 33% overlapping speech from the experimenter were not analyzed; in most cases the researcher elicited a second production during testing, which was the one included in the analysis;
- in addition to the target words, bound, but not free morphemes produced by the child were coded (e.g., ‘washer machine’ and ‘sunglasses’ vs. ‘glasses,’ but not ‘catch *it*’ or ‘*the* cage’);
- the following allophonic variations were not counted as errors:
 - unreleased final stops or glottal release (e.g., /'perɪfʊt/ vs. /'perɪfʊ/)
 - velar stop insertion following a velar nasal (e.g., /'sprɪŋg/)
 - voiced fricative versus affricate production in final position (e.g., /gə'radʒ/ vs. /gə'raʒ/)
 - nasal assimilation substitutions (e.g., /'sænwɪtʃ/ vs. /'sæmwɪtʃ/)
 - partial devoicing of syllable-final voiced consonants (e.g., /'nuzpepər/ vs. /'nuspepər/);
- the allophonic variation –ing reduction (e.g., ‘swimin’) was counted in error;
- phonemes omitted due to rapid speech were counted as inaccurate;
- questionable phonemes that were obscured by noise or recording quality were counted as accurate (Preston & Edwards, 2010);
- vowels were excluded in the analysis, as is standard in the literature on SSD (e.g. Shriberg et al., 1997).

Accuracy judgments were determined using both perceptual and acoustic information in PRAAT (Boersma, & Weenink, 2009). Notes taken on-line during sessions were used for judgment of particular contrasts such as /f/ vs. /θ/, which are difficult to distinguish without visual information. The output produced by running the query PCC in Phon was used to calculate the dependent measure of speech sound accuracy.

Non-word repetition. Procedures were similar to the PNT; templates were created in Phon for each order of the mod-SRT, which were then used to enter broad transcriptions for each item. Transcription and coding conventions followed those outlined by Shriberg and Lohmeier (2008) for the SRT, which included the following:

- manner (stops/nasals) and place (bilabial/alveolar) for consonants were broadly transcribed;
- voicing and lengthening were not counted as errors;
- epentheses were annotated, but not counted as errors;
- vowel substitutions were transcribed, but did not factor into accuracy.

Two other conventions were added based on criteria in the non-word repetition literature (Stokes & Klee, 2009; Preston & Edwards, 2010):

- any stimulus that was replayed for the child due to inattention or any other reason was excluded from the total (< 1% of the data);
- questionable phonemes that were obscured by noise or recording quality were counted as accurate to give the child the benefit of the doubt.

According to the above coding conventions, each consonant in the modified SRT was perceptually judged for accuracy using Phon. Questionable items (< 30 total) were

also analyzed using spectrographic data in PRAAT (Boersma, & Weenink, 2009). For one clinical participant who substituted bilabial stops with velar stops in all word positions, scoring was adjusted so that these consistent substitutions were counted as accurate.

After transcription was completed, a query command in Phon generated an itemized list of each participant's performance including two columns necessary for calculating PCC-R: number of accurate consonants produced compared to target and total number of consonants in the target. The Phon output had to be adjusted by hand to correct two errors: 1) epentheses, which are automatically added to the total number of consonants per item in Phon, and 2) sequencing errors, which are counted as correct even if they are produced out of order. Finally a PCC-R score was derived for each participant by dividing total consonants produced accurately by the total number of target consonants.

Speech perception. The SAILS program (Rvachew, 2010) generated output that included both individual responses by item as well as overall percentage correct for each phoneme. However, due to the potential for response bias inherent in speech discrimination tasks, data were analyzed according to principles of signal detection theory (Keating, 2005). Specifically, the proportion of hits versus proportion of false alarms was calculated for each participant's 50 responses across five phonemes. A d-prime score was calculated in Excel with higher values indicating greater sensitivity in the task.

Reliability of speech production measures

A graduate student from the department of Speech-Language-Hearing: Sciences and Disorders at the University of Kansas who was proficient in broad IPA transcription transcribed 15% of each of the two speech production tasks. The second coder used the blind transcription mode in Phon and followed the coding conventions for each task listed in the previous sections.

Picture Naming Task. Reliability procedures described by Preston and Edwards (2010) were followed to select which items would be transcribed; sixty-three consecutive items (in which the start item was randomly determined) across 20 randomly selected participants (six from the SSD group) were selected for transcription by the second coder, yielding a total of 4,898 consonants. Segments that agreed in place, manner and voicing were counted as agreements (Tyler, Williams, & Lewis, 2006). Inter-rater reliability was calculated using Cohen's kappa. Values greater than .80 are considered good reliability for behavioral data (Wood, 2007); however, typical broad transcription reliability is 85% agreement (Cummings & Barlow, 2011; Shriberg & Lof, 1991).

Overall reliability for the PNT was .825 (Pearson's correlation $r = .827$). Closer inspection revealed that three of the clinical transcripts yielded significantly lower agreement between coders. Those transcripts were reviewed using consensus procedures described by Shriberg, Kwiatkowski and Hoffman (1984). Across the three transcripts, 61 of the 63 disagreements were resolved in favor of the first coder's transcription. Of the remaining 17 transcripts, the mean number of disagreements per participant was 6.2 (range 1-13). Analyses were run a second time on 17 transcripts yielding a Cohen's

kappa of .856 (Pearson's correlation $r = .856$), which is considered an acceptable level of agreement.

Modified SRT. Similar randomization procedures were used to conduct reliability for the modified SRT. Fifteen percent of the data consisted of 14 consecutive items across 20 participants (six from the SSD group) for a total of 800 consonants. Inter-rater reliability for both Cohen's kappa and Pearson's correlation were .996.

Data analysis

All statistical analyses were run using SPSS version 21. Multifactorial ANOVAs and independent group t-tests were used to explore group differences in task performance. Correlational analyses, hierarchical regression, and ANCOVAs were used to examine the relationships between speech sound accuracy, EF task performance, and non-word repetition performance. Alpha was set at $< .05$.

Chapter 3: Results

Participant analyses

Age and gender comparison. SSD and TD groups were well-matched on several dimensions. There were no significant differences in age between the groups (SSD mean in months: 60.5 (SD = 6.5); TD: 59.8 (SD = 6.3); $t(63) = .381$, $p = .704$). Similarly there was no between-group difference in gender distribution (SSD: 9 female, 11 male; TD: 20 female, 25 male; Fisher's exact test, $p = 1.0$). The ratio of males to females with SSD is reported in the literature as 2.75:1 (Shriberg & Kwiatkowski, 1994). Though the gender ratio in this study differed, there was no statistically significant difference in overall proportions between the clinical population in this study and what is described in the literature (Fisher's exact test, $p = .312$).

Demographic comparisons. The two participant groups were equally matched on key developmental milestones such as age of first words, age of walking, positive histories of ear infections, surgeries or illnesses involving mechanisms of speech or hearing, and history of feeding or swallowing problems. Additionally there was no significant between-group difference in the numbers of hours spent per week in structured educational programs including pre-kindergarten and kindergarten ($t(57) = .101$, $p = .920$).

The groups differed, however, on two significant variables: maternal education and familial history of speech, language, or cognitive-communication problems. Mothers of the SSD group generally completed a college degree, whereas more mothers of the TD group completed a master's degree or higher, a difference that reached statistical

significance ($t(56) = 2.356, p = .022$). Several mothers did not complete the questions regarding education, thus the lower degrees of freedom in this statistic.

The proportion of children in the SSD group with a family history of speech, language, reading or cognitive-communication disorder was significantly larger than in the TD group ($t(63) = 4.029, p < .001$). Seventy-nine percent of children with SSD (15 out of 19 respondents) had a first- or second-degree relative with communication problems, while only 31% of parents of TD children (14 out of 45 respondents) reported a similar positive family history. These findings are very much in agreement with current research indicating that speech and language disorders are heritable (Lewis et al., 2006; Stein et al., 2010).

Standardized test performance comparison. Groups were matched on all standardized eligibility criteria as intended. There was a substantial difference in percentile ranks on the GFTA-2 ($t(63) = 24.495, p < .001$) but no statistically significant difference on language and cognitive measures. The TD group did outperform the SSD group on the receptive vocabulary measure (PPVT-4), but this difference did not reach statistical significance ($t(63) = 1.904, p = .061$). Independent samples t-test results comparing percentiles across groups are listed in Table 4.

Table 4. Statistical comparisons on standardized tests

	GFTA-2	CELF-P	PPVT-4	KBIT-2
SSD mean	17 (8.8)	71 (16.0)	79 (14.2)	68 (18.2)
TD mean	80 (10.0)	71 (20.3)	86 (15.3)	71 (20.6)
Statistic	$t(63) = 24.495^*$	$t(63) = .082$	$t(63) = 1.904$	$t(63) = .567$
Significance	$p < .001$	$p = .935$	$p = .061$	$p = .573$

Means are by percentile. Standard deviations are listed in parentheses. * indicates p-value is $< .05$.

The difference in receptive vocabulary might have been related to the slightly higher education attained by mothers of the TD group (as discussed in participant demographics); there was a moderate correlation between maternal level of education and children's PPVT-4 percentile ($r = .317, p = .015$). This relationship is consistent with research suggesting that maternal education is predictive of children's receptive vocabulary at age four (Taylor, Christensen, Lawrence, Mitrou & Zubrick, 2013).

Spontaneous language comparison. Two expressive language measures, DSS and VOC-D, were compared between groups using independent samples t-tests. Data from three participants (two SSD, one TD) were excluded because the child did not produce at least 40 greater than single-word utterances and/or speech production was too unintelligible to accurately transcribe. As can be seen in Table 5, the groups were equally matched in syntactic usage ($t(60) = .515, p = .608$), but not in lexical diversity ($t(60) = 2.243, p = .029$). Overall, the typically-developing speech group demonstrated a greater mean type-token ratio than the children with SSD.

Table 5. Statistical comparisons of spontaneous expressive language measures

	DSS (syntactic complexity)	VOC-D (vocabulary diversity)
SSD mean	8.19 (1.58)	59.42 (14.81)
TD mean	7.94 (1.77)	70.92 (19.04)
Statistic	$t(60) = .515$	$t(60) = 2.243^*$
Significance (p < .05)	$p = .608$	$p = .029$

Means are derived from CLAN analysis values. Standard deviations are listed in parentheses. * indicates p-value is < .05.

Speech sound accuracy measures

As described in the methods, the GFTA-2 was used to determine group assignment, while the PNT was used as the primary measure of speech sound accuracy. A Pearson's correlation was run to examine the consistency between these two measures. Results demonstrated a high correlation between PNT scores and GFTA standard scores ($r = .933, p < .001$). The lack of perfect correlation can best be explained by differences in scoring systems used. Specifically, the GFTA-2 was scored using Percent Consonants Correct, a stricter system of speech sound classification. The PNT, on the other hand, was scored using PCC-Revised (PCC-R), a system in which distortions are counted as accurate productions rather than errors. This discrepancy in scoring systems also explains why one child initially met the eligibility requirements for the clinical group, but was later excluded according to results of the PNT. Table 6 lists data from the PNT. Similar to the GFTA-2, results of an independent samples t-test showed a significant difference between groups in PCC-R scores ($t(63) = 10.508, p < .001$).

Table 6. PCC-R for PNT by group

	Mean	Std deviation	Range
SSD	69.3%	10.48	48.5 – 84.0%
TD	94.4%	2.98	85.3 – 98.9%

Two additional analyses were conducted to examine whether PNT scores were significantly influenced by age or gender. This step was not necessary for the GFTA-2 because it is standardized, whereas either or both of these variables could potentially confound the PNT results. A Pearson's correlation confirmed that age in months and PCC-R score were significantly correlated ($r = .378, p = .002$). Although boys achieved slightly lower PCC-R scores than did girls (86.1% versus 87.3%), this difference was not significant as indicated by an independent samples t-test ($t(63) = .380, p = .705$). As a result of these analyses, age in months was statistically controlled in all analyses involving PCC-R for the PNT, but gender was not.

Executive Function performance

The first few research questions asked whether there were between-group differences in EF skills that might help explain why children with SSD require explicit instruction to correct their errors. The following sections examine group performance for each EF domain and research hypothesis. Table 7 is a list of scores and analyses by group and task.

Table 7. EF task raw scores and statistics (with age as a covariate)

Task	SSD	TD	Statistic
	Mean (SD) Range	Mean (SD) Range	Significance
Flowers (accuracy) Max score: 20	10.9 (4.73) 2 - 17	12.0 (4.6) 4 - 20	$F(1, 62) = 1.425$ $p = .237$
Flowers (RT)	971.13(148) 971-1229	967.9 (130) 592-1187	$F(1, 62) = .011$ $p = .917$
Modified DNS Max score: 20	11.8 (4.4) 0 - 19	12.7 (4.2) 2 - 20	$F(1, 62) = .806$ $p = .373$
Hearts and Flowers (accuracy) Max score: 20	10.9 (4.5) 1 - 20	11.6 (3.6) 5 - 19	$F(1, 62) = .644$ $p = .425$
Hearts and Flowers (RT)	1161.5(307) 655-1795	1191.2 (269) 582-1672	$F(1, 62) = .245$ $p = .622$
FIST Max score: 12	6.2 (3.4) 0 - 11	7.6 (3.2) 0 - 12	$F(1, 62) = 4.167^*$ $p = .045$
Forward digit span	3.8 (.7) 3 - 5	4.2 (.8) 3 - 6	$F(1, 62) = 6.110^*$ $p = .016$
Backward digit span	1.9 (1.1) 0 - 3	2.2 (1.1) 0 - 4	$F(1, 62) = .926$ $p = .340$
Animal span Max score: 37	27.5 (4.4) 17 - 35	29.0 (4.0) 20 - 37	$F(1, 62) = 2.426$ $p = .124$

Accuracy data are reported in raw scores. Standard deviations are in parentheses. RTs are in milliseconds. * indicates p-value <.05.

Hypothesis 1: Differences in inhibitory control. The first hypothesis predicted that children with cognitive-linguistic SSD would perform more poorly on inhibitory control tasks than their typical-speech peers. Performance on the Flowers task-

both accuracy and RT- and the modified DNS were examined to test this hypothesis.

Three separate ANCOVAs were run using age in months as a covariate. The rationale for co-varying age for all EF tasks was to statistically control for differences based solely on maturation (Chevalier, Huber, Wiebe & Andrews Espy, 2013). As has been shown in the executive function literature, young four-year-olds perform differently on EF tasks than older five-year-olds (Carlson, 2005; Diamond et al., 2002).

Although the SSD group underperformed their TD peers on both tasks, there were no statistically significant differences between groups on the Flowers task for accuracy ($F(1, 62) = 1.425, p = .237$) or reaction times ($F(1, 62) = .011, p = .917$). Similarly, performance differences on the modified DNS were not significant ($F(1, 62) = .806, p = .373$). These results suggest that inhibitory control may not contribute to children's difficulty transitioning to an adult-like speech sound system.

Hypothesis 2: Differences in cognitive flexibility. The second hypothesis predicted that children with SSD would demonstrate less cognitive flexibility than TD children. The Hearts and flowers task and FIST were used to compare groups in this EF domain. As with the inhibitory control analyses, three ANCOVAs were run using age in months as a covariate. Although children with SSD were less accurate on their performance on the Hearts and flowers task, this result was not statistically significant ($F(1, 62) = .644, p = .425$). Interestingly, the SSD group was faster to respond than their TD peers, although this result was not statistically significant ($F(1, 62) = .245, p = .622$). However, it should be noted that accuracy for all participants in this task was close to chance performance (chance = 10; SSD mean = 10.85; TD mean = 11.64); reaction times likely had little to do with efficiency of processing.

In contrast to the Hearts and flowers task, group comparisons for the FIST were statistically significant ($F(1, 62) = 4.167, p = .045, \eta^2 = .063$); children with SSD demonstrated lower accuracy than their TD peers in shifting from one response to another. These findings are in agreement with Dodd and colleagues (Crosbie et al., 2009), even though responses were scored differently between the studies. Specifically, Crosbie et al. counted children's accuracy of both first and second choices under their criteria for "rule abstraction," while this study only counted second choices when the first selection was accurate. Notably there was no significant between-group difference in accuracy for participants' first selections ($F(1, 62) = 1.195, p = .279$). Overall, these results suggest that children with SSD may demonstrate less cognitive flexibility, at least on particular tasks, than children with typical speech development.

Hypothesis 3: Differences in phonological memory. The third hypothesis predicted that children with SSD would have poorer phonological memories than their typically-developing peers. To test this hypothesis, ANCOVAs with age as a covariate were run on each of the two digit span tasks and the animal span task. Again, children with SSD underperformed their peers on all tasks of phonological memory. There was only a small difference, however, between groups on the backward digit span task ($F(1, 62) = .926, p = .340$), which proved difficult for most participants. Twenty percent of the children with SSD and 15% of the TD group were unable to mentally manipulate even two digits. Differences in performance on the animal span task were also not statistically significant ($F(1, 62) = 2.426, p = .124$).

The forward digit span task, on the other hand, showed a statistically significant difference between groups ($F(1, 62) = 6.110, p = .016, \eta^2 = .090$). This result is in

agreement with a handful of studies showing lower digit span performance in children with SSD (Lewis et al., 2011; Tkach et al., 2011). Taken together, these results suggest that deficits in short-term memory capacity, rather than mental manipulation, may be seen in this population.

Hypothesis 4: The relationships between speech sound accuracy and EF tasks. Perhaps even more informative than group differences in EF task performance is an analysis of the relationships between each of the three core EFs and speech sound development. The fourth hypothesis predicted that tasks requiring cognitive flexibility and inhibitory control would be more strongly related to speech sound accuracy than would phonological memory. It was also predicted that phonological memory would relate more strongly to performance in other EF tasks than to speech sound accuracy *per se*.

Speech sound accuracy and EF tasks. To test the first part of this hypothesis, Pearson's partial correlations, using age as a control variable, were run between each of the EF tasks and the PNT PCC-R. This analysis used z-scores for the EF tasks rather than raw scores. The use of z-scores was selected to standardize each measure for the analysis; as would be expected, the relationship between raw and z-scores for each task showed a perfect correlation ($r = 1.0$).

Results from this correlational analysis (see Table 8) indicated that only the forward digit span task was related to speech production accuracy ($r = .388, p = .002$). The correlation was positive, indicating that children who accurately repeated more digits had greater speech sound accuracy. Although the correlation with backward digit span approached significance ($r = .328, p = .056$), an additional correlation in which forward

digit span was partialled out effectively made this relationship disappear ($r = .128, p = .312$). These results suggest that memory capacity versus manipulation was the more related process. No other correlations between EF tasks and the PNT were statistically significant.

Table 8. Partial correlation matrix between EF tasks and PNT (with age as a control variable)

	PNT	Forw	Back	Anim	DNS	Flow	FIST	H&F
PNT <i>r</i>	1.000	.388**	.240	.099	.089	.100	.161	.041
<i>p</i>	.	.002	.056	.436	.486	.431	.204	.745
Forw <i>r</i>	.388**	1.000	.328**	-.059	.087	-.078	.322**	.075
<i>p</i>	.002	.	.008	.643	.493	.542	.009	.558
Back <i>r</i>	.240	.328**	1.000	.191	.056	-.176	.125	.092
<i>p</i>	.056	.008	.	.130	.658	.165	.324	.468
Anim <i>r</i>	.099	-.059	.292	1.000	.075	.429**	.123	.196
<i>p</i>	.436	.643	.130	.	.555	.000	.332	.120
DNS <i>r</i>	.089	.087	.056	.075	1.000	-.048	.074	.192
<i>p</i>	.486	.493	.658	.555	.	.707	.559	.129
Flow <i>r</i>	.100	-.078	-.176	.429**	-.048	1.000	.151	.328**
<i>p</i>	.431	.542	.165	.000	.707	.	.233	.008
FIST <i>r</i>	.161	.322**	.125	.123	.074	.151	1.000	.026
<i>p</i>	.204	.009	.324	.332	.559	.233	.	.841
H&F <i>r</i>	.041	-.075	-.092	.196	.192	.328**	.026	1.000
<i>p</i>	.745	.558	.468	.120	.129	.008	.841	.

** indicates p-value is < .01. Forw = Forward digit span, Back = Backward digit span, Anim = Animal span, Flow = Flowers task, H&F = Heart and flowers task.

A secondary analysis using hierarchical regression further explored this relationship. The goal of a regression analysis was to examine whether any of the EF tasks that significantly differed between groups was predictive of speech sound accuracy. A step-wise regression model was run in which PNT functioned as the dependent variable and two EF tasks- forward digit span and the FIST- were entered into the second

step of the model. Age in months was entered as the first step in the model to ensure that any variability from this confounding variable was removed in advance.

Results from the second step of the model were significant ($F(3,61) = 4.016, p = .011$) and in agreement with the correlational analysis, indicating that performance on the forward digit span was the only task that predicted speech sound accuracy (see Table 9). According to the model, age, forward digit span, and the FIST together accounted for 16.5% (r -squared) of the variability in speech sound production (12.4% adjusted r -squared). Taken together, these results suggest that phonological memory capacity may be the only core EF skill specifically implicated in speech sound development.

Table 9. Regression analyses of variables used to predict PNT scores

Variable	<i>B</i> (SE)	β	<i>t</i>	<i>p</i>
Age in months	.001 (.003)	.037	.289	.773
Forward digit span	.050 (.017)	.379	3.014**	.004
FIST	.006 (.018)	.043	.321	.749

** indicates p value $< .01$.

EF task relationships. It was also predicted that phonological memory would relate more to performance on the other tasks than it would to speech sound development. This question is somewhat less relevant considering that phonological memory, as measured by forward digit span, was found to be the primary EF component that related to speech sound accuracy. The relationship among tasks is nevertheless of considerable theoretical interest.

As seen in the correlation matrix in Table 8, forward digit span was moderately correlated with performance on both backward digit span ($r = .328, p = .008$) and the FIST ($r = .322, p = .009$). There were other notable relationships in the matrix; the Flowers task was moderately correlated with both the Hearts and flowers task ($r = .328, p = .008$) and the animal span task ($r = .429, p < .001$). While the correlations between yoked tasks are not particularly surprising (i.e., digit span tasks and Hearts and Flowers tasks), the relationships between all EF tasks will be explored further in the next section.

Exploratory analysis 1: Analysis of EF constructs. One potentially arguable assumption in this study was whether each task was correctly assigned to a specific EF construct. Not only do many researchers suggest that EFs rarely work in isolation (Davidson et al., 2006; Diamond, 2013), but the tasks assigned to each construct could possibly have been inaccurate. An additional consideration is that individuals- particularly children- might differ on how they approach a task, which could effect which EF is primarily being employed (Dauveir et al, 2012; Ramscar et al., 2013).

Given the nature of these concerns, along with results from the task correlations in the previous section, a factor analysis was conducted. The purpose of this analysis was to examine whether tasks fell under specific constructs as designed or whether they might group according to other latent variables. Results from the component matrix using varimax rotation with kaiser normalization (Table 10) showed a two-factor solution with the first factor explaining 28.7% and the second explaining 21.7% of the variance. Similar weights were obtained when children from the SSD group were excluded.

Table 10. Factor analysis of EF tasks with two-factor solution

	Component	
	1	2
Forward span	-.032	.797
Backward span	-.120	.707
Animal span	.711	.183
DNS	.335	.275
Flowers	.829	-.048
FIST	.407	.588
Hearts & flowers	.656	-.179

This analysis agreed with results from the correlation matrix: digit span tasks and FIST loaded onto one construct, while the hearts and flowers tasks, animal span task, and DNS- although weakly- loaded onto a second construct. Implications for these results will be considered in the discussion section.

Exploratory analysis 2: Speech perception. Speech sound discrimination was included in the test battery as an exploratory variable. As described in the methods section, discrimination was quantified as a composite score across five phonemes. D-prime values were derived in order to provide a more sensitive measure of speech sound discrimination by decreasing effects of response bias.

The first question of interest was whether d-prime values differed between groups; a significant difference would support evidence indicating that some children with SSD have generalized deficits in speech perception (Rvachew & Brosseau-Lapre,

2013; Rvachew et al., 1999). Results showed that although d-prime values were lower in SSD children as compared to controls [SSD mean = 1.180 (SD = .681); TD mean = 1.429 (SD = .835)], an independent samples t-test showed no significant difference between groups ($t(63) = 1.171, p = .246$). In a secondary quantitative analysis, d-prime values across all participants were divided into quartiles in order to compare the composition of the lowest 25% to the highest 25%. The upper quartile consisted of 12 TD versus four SSD children, while the bottom quartile consisted of 11 TD versus six SSD children; this was not a statistically significant difference (Fisher's exact test, $p = .708$).

The second question that was examined using this exploratory variable was whether speech sound discrimination was related to speech sound accuracy, as would be predicted by the literature (Rvachew et al., 2004; Rvachew et al., 1999). Though results from a Pearson's correlation approached significance ($r = .210, p = .094$), the relationship disappeared when age was partialled out ($r = .175, p = .167$). These analyses either question the validity of the speech perception measure and/or cast doubt on the relationship between generalized speech perception and production at least in this group of children. In either case, this variable was not used in other analyses.

Non-word repetition

The second part of this study explored non-word repetition in children with SSD using modifications from a protocol introduced by Shriberg and colleagues (2009). Data were analyzed to examine a number of research questions including performance differences between SSD and TD children, effects of stress and number of syllables, and the relationship between non-word repetition and EF task performance. Table 11 lists the

data, although prior to running all analyses percentages were converted to arcsin values in order to normalize the data.

Table 11. Modified SRT accuracy percentages and statistics by group (with continuity condition as a covariate)

	SSD	TD	Statistic
	Mean (SD) Range	Mean (SD) Range	Significance
PCC-R overall	70.8 (14.7) 41.8-89.5	81.7 (6.4) 44.2-97.7	$F(1, 62) = 8.046^{**}$ $p = .006$
PCC-R equal stress	67.4 (16.0) 37.2-95.3	76.4 (12.6) 48.8-100	$F(1, 62) = 7.448^{**}$ $p = .008$
PCC-R stressed	70.6 (14.8) 46.5-93.0	78.8 (12.8) 39.5-97.7	$F(1, 62) = 6.348^{*}$ $p = .014$
PCC-R 2-syllable	87.5 (14.2) 56.3-100	96.8 (5.3) 81.3-100	$F(1, 62) = 14.794^{**}$ $p < .001$
PCC-R 3-syllables	80.4 (16.2) 55.6-100	90.0 (12.3) 50.0-100	$F(1, 62) = 8.470^{**}$ $p = .005$
PCC-R > 3 syllables	59.5 (16.7) 28.8-84.6	67.5 (15.7) 25.0-96.2	$F(1, 62) = 4.251^{*}$ $p = .043$

Values reflect percent correct. Standard deviations are in parentheses. * indicates p-value is <.05. ** indicates p-value is <.01.

Confounding variables. Prior to running any between-group analyses, it was important to explore the effects of any extraneous variables that might inadvertently

confound results. These variables included age, gender, and task administration. Each of these variables will be examined in turn.

Age was a significant factor in both PNT accuracy and some EF tasks, while gender was not; similar analyses were conducted with these two variables to examine their effects on the modified-SRT. Results from a Pearson's correlation between age in months and PCC-R for the mod-SRT indicated that there was no significant relationship ($r = .208, p = .096$). The effect of gender on mod-SRT performance was examined using an independent samples t-test. Although boys had slightly lower accuracy than girls in this task (74.8% and 75.2%, respectively), the difference was not significant ($t(63) = .094, p = .926$). Based on these results, neither age nor gender was factored out in the experimental analyses.

The final potentially confounding variable that was examined was a difference in the way in which the task was administered. As described in the methods section, the first 25 participants (8 SSD, 17 TD) performed the 26-item task without breaks, while the remaining children (12 SSD, 28 TD) were given a puzzle piece after every other trial as a reinforcing activity. Though there was no significant difference in the proportion of each group by condition (Fisher's exact test, $p = 1.00$), it was important to examine whether this variation in protocol affected task performance. It was reasoned that task performance could be affected in one of two ways: 1) the puzzle activity could improve performance by decreasing carryover effects between like items, or 2) it could negatively impact performance by distracting children from the task.

Results from an independent samples t-test showed that there was a significant difference in performance between children by condition ($t(63) = 3.685, p < .001$).

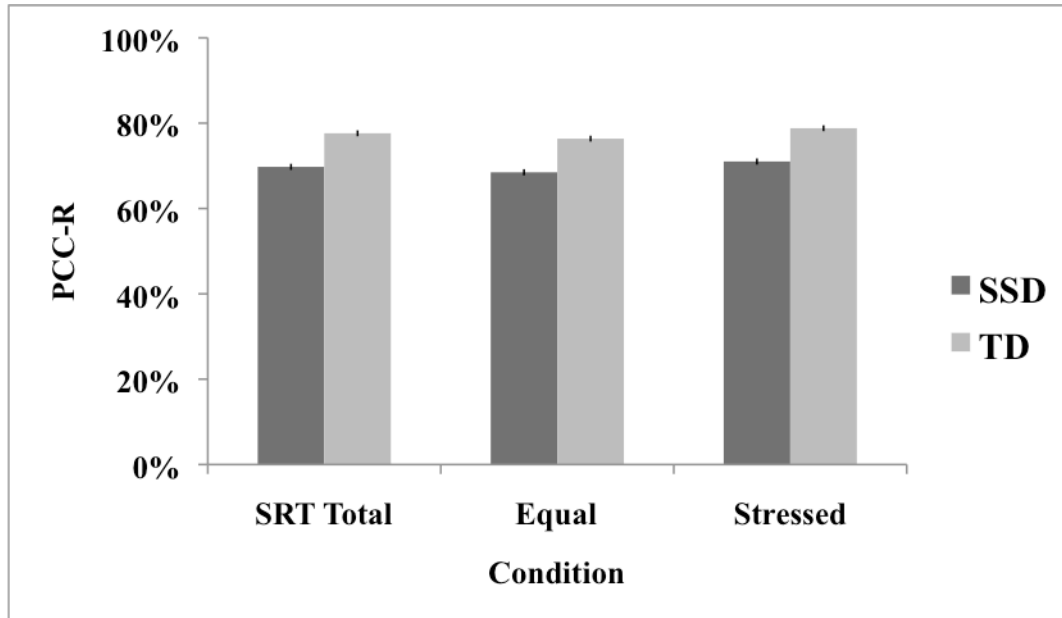
Specifically, participants who did the puzzle activity had lower mod-SRT accuracy (mean = 70.8%, SD = 14.7) than children who performed the task without interruption (mean = 81.7%, SD = 6.4). To test whether this effect was not actually caused by differences in speech sound accuracy (e.g., children in the puzzle condition had lower speech sound accuracy in general), a t-test was run comparing PNT scores by mod-SRT condition. The results of this analysis were not statistically significant ($t(63) = .612, p = .543$), indicating that the effect of condition was due to task administration rather than participant characteristics. Children were likely distracted by the puzzle reinforcer, which in turn affected repetition accuracy. Because of this result, the condition of continuity was used as a control variable in all subsequent analyses.

Hypothesis 5: Differences in mod-SRT performance. The fifth hypothesis predicted that children with SSD would perform more poorly on the mod-SRT than TD children due to deficits in underlying cognitive processes recruited in NWR tasks. In addition, it was predicted that these differences would remain even if age and language abilities were controlled. Although a significant relationship between age and mod-SRT performance was already discounted, the effects of both age and language factors were examined. Finally, it was predicted that performance in the mod-SRT would positively relate to speech sound accuracy.

In response to the primary hypothesis, an ANCOVA was run comparing group performance (SSD vs. TD) on the mod-SRT using continuity condition (puzzle vs. no puzzle) as a control variable. Results from this analysis confirmed a significant difference in accuracy for the mod-SRT ($F(1, 62) = 7.561, p = .008, \eta^2 = .109$), demonstrating that

children in the clinical group were considerably less accurate than their TD peers. These data are summarized in Table 11 and Figure 7.

Figure 7. Results of the mod-SRT by group and task condition



Next, a series of analyses explored whether controlling age or language skills affected group differences on the mod-SRT. Several ANCOVAs were run using two covariates: 1) continuity condition, and 2) age in months, vocabulary, or language variables. Results demonstrated that significant differences between groups remained after controlling for age ($F(1,61) = 8.363, p = .005, \eta^2 = .121$), receptive language scores (CELF; $F(1,61) = 4.527, p = .037, \eta^2 = .069$), receptive vocabulary scores (PPVT-4; $F(1,61) = 7.375, p = .009, \eta^2 = .108$), and expressive language skills (DSS; $F(1,58) = 8.101, p = .006, \eta^2 = .117$). However, the between-group difference on the mod-SRT disappeared when lexical diversity was used as a control variable (VOC-D; $F(1,58) =$

2.678, $p = .107$). The relationship between mod-SRT performance and lexical diversity in spontaneous speech was further substantiated by a rather large effect size ($r = .442$, $p < .001$). This relationship will be addressed further in the discussion.

The final question in this section examined whether performance on the mod-SRT was related to speech sound accuracy. A partial Pearson's correlation using continuity condition as a control variable showed a moderate correlation between accuracy on the mod-SRT and the PNT ($r = .386$, $p = .002$; see Table 12 for the complete matrix). This positive correlation indicated that children with greater accuracy on the NWR task also had better speech sound accuracy. Furthermore, almost 14.9% of the variance on the mod-SRT was accounted for by speech sound accuracy. Because the SRT was specifically designed for children with SSD and consists only of early-acquired consonants, the r -squared value cannot be attributed to articulation accuracy. Rather, some underlying mechanism, such as encoding or phonological memory, is implicated in this relationship.

Table 12. Partial correlation matrix for mod-SRT (controlling for continuity condition)

		PNT	SRT	equal	stressed	2-syll	3-syll	>3-syll
PNT	<i>r</i>	1.000	.386	.385	.331	.465	.384	.300
	<i>p</i>	.	.002	.002	.007	.000	.002	.016
SRT	<i>r</i>	.386	1.000	.934	.933	.565	.707	.975
	<i>p</i>	.002	.	.000	.000	.000	.000	.000
Equal	<i>r</i>	.385	.934	1.000	.745	.567	.666	.909
	<i>p</i>	.002	.000	.	.000	.000	.000	.000
Stressed	<i>r</i>	.331	.933	.745	1.000	.487	.646	.913
	<i>p</i>	.009	.000	.000	.	.000	.000	.000
2-syll	<i>r</i>	.465	.565	.567	.487	1.000	.459	.430
	<i>p</i>	.000	.000	.000	.000	.	.000	.000
3-syll	<i>r</i>	.384	.707	.666	.646	.459	1.000	.580
	<i>p</i>	.001	.000	.000	.000	.000	.	.000
>3-syll	<i>r</i>	.300	.975	.904	.909	.430	.580	1.000
	<i>p</i>	.016	.000	.000	.000	.000	.000	.

Analyses were run using arsin values. All correlations were significant at $< .01$ significance. Equal = original SRT items, Stressed = items with word stress, Syll = syllable.

Hypothesis 6: Effect of word stress. As discussed previously, the purpose of modifying the original SRT was to examine whether the addition of word stress would affect performance in all participants or by group. It was hypothesized that across groups, participants would demonstrate higher accuracy on modified SRT items with prosodic variation than on items with equal stress (i.e., original SRT items). It was also predicted that children with SSD would gain equal benefit from word stress as their TD peers. The second part of this hypothesis predicted that children in the SSD group would demonstrate lower accuracy on unstressed as compared to stressed syllables.

For the first analysis comparing effects of prosodic stress across groups, a 2x2 repeated-measures ANCOVA (within-group: stressed vs. equal stress, between-group:

SSD vs. TD) was run using continuity condition as a covariate. Results (see Figure 7) again showed the main effect of group on overall task performance ($F(1, 62) = 7.391, p = .008, \eta^2 = .107$), but no main effect of word stress ($F(1, 62) = .304, p = .583$) or interaction term ($F(1, 62) = .071, p = .791$). In other words, the presence of stress cues for SRT items did not improve accuracy across participants as predicted, nor did it differentially affect participants by group. These null results were somewhat surprising based on evidence in the literature (Morgan et al., 2013, but see Gupta, Lipinski, Abbs & Lin, 2005) and will be discussed further.

Next, it was examined whether the trochaic stress pattern, which is more common in English, would result in greater accuracy than items with iambic stress. It was also explored whether there would be a difference in how groups responded to these items. This analysis was run on the 13 stressed items in the modified condition. A repeated measures 2 x 2 ANCOVA (within-group: trochaic vs. iambic, between-group: SSD vs. TD; covariate: continuity condition) revealed a significant main effect between groups ($F(1, 62) = 4.882, p = .031, \eta^2 = .073$), but no effect of stress pattern ($F(1, 62) = 2.086, p = .154$) and no significant interaction ($F(1, 62) = .691, p = .409$). Surprisingly, as can be seen from the data in Table 13, accuracy for both groups was better for items with iambic than trochaic stress pattern. The difference between groups was consistent with overall mod-SRT performance, but the non-significant effect of stress pattern was contrary to prior evidence showing considerably greater accuracy for trochaic over iambic stress in real words (Gerken & McGregor, 1998).

The final analysis with regard to stress effects in non-word repetition examined whether consonants in stressed syllables were produced with greater accuracy than in

unstressed syllables and whether this effect would be similar for both groups. The proportions of accurate consonants in stressed and unstressed syllables were calculated for each group (see Table 13), converted to arcsin values, and then analyzed using a repeated measures 2 x 2 ANCOVA (within-group: stressed vs. unstressed syllables; between-group: SSD vs. TD; covariate: continuity condition). Results showed two main effects between group ($F(1, 62) = 5.840, p = .019, \eta^2 = .086$) and stress ($F(1, 62) = 4.564, p = .037, \eta^2 = .069$), but no interaction ($F(1, 62) = .230, p = .633$). These results indicated that participants produced consonants in stressed syllables more accurately than in unstressed syllables. Importantly, children in the SSD group were no more vulnerable to unstressed syllable errors than their TD peers.

Table 13. Mean proportions of accurate consonants and between group statistics by stress pattern and stressed syllables

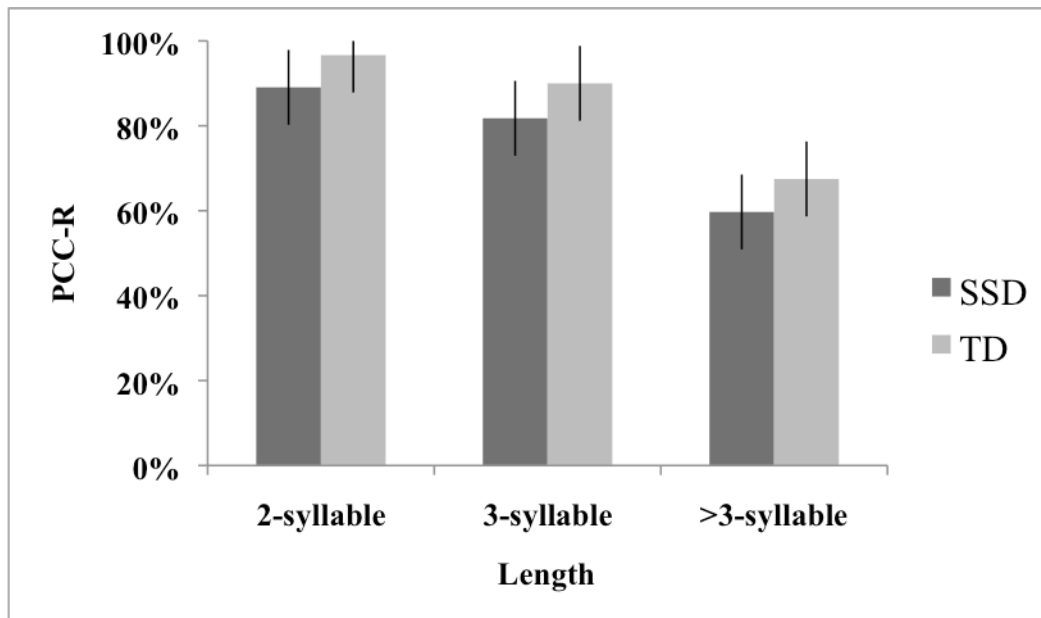
	SSD		TD		Statistic
	Mean (SD)	Range	Mean (SD)	Range	p-value
Trochaic	.678 (.14)	.40-.95	.737 (.14)	.45-.90	$F(1,62) = 3.253$ $p = .076$
Iambic	.743 (.18)	.48-1.0	.832 (.15)	.30-1.0	$F(1,62) = 4.091^*$ $p = .047$
Stressed syllables	.730 (.15)	.47-1.0	.814 (.13)	.47-1.0	$F(1,62) = 5.025^*$ $p = .029$
Unstressed syllables	.682 (.18)	.29-.93	.761 (.15)	.32-.96	$F(1,62) = 4.501^*$ $p = .038$

Analyses were run using arcsin values. * indicates p-value is < .05.

Hypothesis 7: Effect of syllable length. The research question of interest in this section was the effect of word- or syllable-length on non-word repetition accuracy. Specifically, it was hypothesized that accuracy would decrease as item length increased. In addition, based on previous research of this effect, it was predicted that there would be no differential group response (Lewis et al., 2011; Munson et al., 2005; Preston & Edwards, 2007; Roy & Chiat, 2004; Shriberg et al., 2009).

To address this question, a 3 x 2 repeated-measures ANCOVA (within-group: 2-syllables, 3-syllables, 4-5-syllables, between-group: SSD vs. TD) was conducted using continuity condition as a covariate (see Figure 8). There were main effects of group ($F(1, 62) = 14.175, p < .001, \eta^2 = .186$) and length ($F(1, 62) = 23.049, p < .001, \eta^2 = .271$). In addition, there was no interaction ($F(1, 62) = 1.734, p = .181$), meaning that groups showed similar patterns with regard to length. Follow-on paired t-tests demonstrated significant differences between all lengths: 2-syllable versus 3-syllables items ($t(64) = 4.584, p < .001$), 3-syllable versus 4-5-syllable items ($t(64) = 13.926, p < .001$), and 2-syllable to 4-5-syllables items ($t(64) = 18.588, p < .001$). These results confirm the predictions that increasing syllable-string length worsens repetition accuracy and that children with SSD are less accurate at all non-word lengths than TD children.

Figure 8. Results of the mod-SRT by group and stimulus length



One additional concern raised by Shriberg et al. (2009) and other researchers interested in non-word repetition tasks, is that shorter items (i.e., 2-syllable stimuli) are commonly in error among clinical groups. If phonological memory processes are what is measured in this task, it would be predicted that shorter items would not differ between groups; however, the data from this study are in line with patterns reported in the literature, which run counter to this prediction. Children from the SSD group were on average 88% accurate with two-syllable items as compared to their TD peers' 97% accuracy. Results from an ANCOVA comparing group performance differences on 2-syllables items (using arcsin transformations) with continuity condition as a covariate were statistically significant ($F(1, 62) = 14.794, p < .001, \eta^2 = .193$). This issue will be explored further in the next section, which considers both phonological memory tasks (e.g., digit span) and non-word repetition.

Hypothesis 8: Non-word repetition, EFs and speech sound accuracy.

The final set of analyses explored relationships between speech sound accuracy, the modified SRT, and EF task performance. The first question asked whether performance on any of the EF tasks would relate to accuracy on the mod-SRT; it was predicted that only phonological memory tasks would be related. Partial Pearson's correlations using continuity condition and age as controlled variables confirmed this prediction. Only forward digit span ($r = .285, p = .024$) was positively correlated with SRT accuracy (see Table 14 for correlation matrix).

Table 14. Correlations between mod-SRT and EF tasks (continuity condition and age in months are partialled out)

	Forw	Back	Anim	DNS	Flow	FIST	H&F
SRT <i>r</i>	*.285	-.019	.234	.180	.174	.050	.122
<i>p</i>	.024	.885	.065	.158	.173	.699	.341

* indicates significance is $< .01$. Forw = Forward digit span, Back = Backward digit span, Anim = Animal span, Flow = Flowers task, H&F = Heart and flowers task.

The next question explored whether the processes underlying performance in mod-SRT and phonological memory tasks independently contributed to speech sound accuracy. This was a question that has been examined with regard to language skills, specifically to SLI (Archibald & Gathercole, 2007), but not to speech skills and SSD. It was hypothesized that the mod-SRT would recruit slightly different processes than the digit span task, and that these differences could be teased apart when related to speech sound accuracy.

For this analysis, a hierarchical regression model was used with PNT PCC-R as the dependent variable. In the first step of the model, age and continuity condition were entered, to ensure that any variation attributed to these variables was removed; results from this step were not significant ($F(2, 62) = .575, p = .566$). The second step of the model, in which forward digit span and mod-SRT were entered as predictors, was statistically significant ($F(4,60) = 4.693, p = .002$; see Table 15 for regression results). In addition, the r -squared value for the model with both predictors was 23.8% (adjusted r -squared 18.8%) in comparison to the earlier regression model in which forward digit span alone accounted for only 16.4% of the variance (13.7% adjusted r -squared) on the PNT. These results indicate that forward digit span and mod-SRT are both unique predictors of speech sound accuracy.

In sum, results suggest that forward digit span and NWR tap into at least partially distinct underlying cognitive processes, both of which are affected in children with SSD. These results might also help explain why children with SSD perform more poorly on 2-syllable non-word repetition than their TD peers because this task requires more than just working memory capacity (although that too is impaired). Further implications will be discussed in the next chapter.

Table 15. Regression analyses of mod-SRT and forward digit span as predictors of PNT scores

Variable	B (SE)	β	t	p
Forward digit span	.055 (.021)	.337	2.650*	.010
Mod-SRT	.004 (.002)	.300	2.323*	.024

* indicates p value < .05

Chapter 4: Discussion

Overview

This study examined whether a difference in core executive functions might help explain why children with cognitive-linguistic SSD need explicit help to correct their speech sound errors. Results demonstrated between-group performance differences in only two tasks: forward digit span, used as a measure of phonological working memory capacity, and the FIST, which is proposed to measure cognitive flexibility. Of these two tasks, only forward digit span was significantly correlated with and predictive of speech sound accuracy. Interestingly, forward and backward digit spans and the FIST were correlated with each other and also factored together, suggesting a similar underlying cognitive component. The modified Day-Night Stroop task (DNS), animal span task and Hearts and Flowers tasks were not found to be related to speech sound accuracy, but they factored together under a different latent variable.

The second part of the study explored a number of questions using a recently published non-word repetition task by Shriberg et al. (2009) that avoids the necessity of correcting for children's articulation errors. The task was designed with two conditions to examine whether the addition of word stress would enhance the sensitivity of the original task. It was found that children with SSD demonstrated lower accuracy overall than TD children on the modified SRT. There was a consistent length effect, such that longer items were repeated less accurately than shorter items, although children with cognitive-linguistic SSD also demonstrated poorer accuracy on two-syllable items than the TD participants. The addition of word stress on the mod-SRT did not affect performance within or between groups, nor did the type of stress (trochaic or iambic) in the stressed

condition. As might be expected, stressed syllables were produced with greater accuracy than unstressed syllables across all participants. Finally, performance on the mod-SRT task, which purportedly measures phonological working memory, was related to both speech sound accuracy and forward digit span. Both forward digit span and the mod-SRT were unique predictors of speech sound accuracy.

Inhibitory control, cognitive flexibility, and cognitive-linguistic SSD

Though it was hypothesized that inhibitory control and cognitive flexibility might be two core EF skills underlying the transition to adult-like speech, the four tasks chosen to measure these constructs were found to be unrelated to speech sound accuracy. One possibility for the null results is that in order to inhibit or flexibly shift to an alternate production, children must first perceive that their production patterns are wrong. Several researchers have suggested that children with cognitive-linguistic SSD do not identify their productions as being mismatched with adult targets (Kornfeld & Goehl, 1974; Shuster, 1998; Strombergsson, Wengelin & House, 2014). If underlying representations are not perceived as requiring self-correction, then core EFs may not be involved in the process of spontaneously transitioning to adult-like phonology.

On the other hand, null results do not necessarily rule out the involvement of these cognitive processes. One alternate explanation is that the selected tasks did not tap into the same underlying skills required in order to self-correct. For instance, skills requiring motor output, such as required by Hearts and Flowers tasks, may be very different from the cognitive skills involved in inhibiting or flexibly shifting between phonological representations. Similarly, the modified DNS likely requires inhibitory

control at the lexico-semantic level, which may be different from inhibitory control at the phonological level. Likewise, shifting between salient features on the FIST might not be a generalizable skill used in other tasks requiring cognitive flexibility, such as overwriting early phonological representations. The concern that EF tasks reflect targeted rather than general constructs is a relevant topic in the EF literature as a whole, and one that has significant implications for future translation to clinical applications (Jaeggi, Muschkuehl, Jonies & Shah, 2011; Shipstead, Redick & Engle, 2012).

Phonological working memory and cognitive-linguistic SSD

Study results showed both group differences and a positive correlation between forward digit span and speech sound accuracy. The relationship with backward digit span approached significance, but not when forward digit span was statistically controlled. This finding is consistent with both neurocognitive and psycholinguistic models in which digit span tasks share some neural networks, but also reflect distinct processes in working memory (Baddeley, 2001; Østby et al., 2011). These findings are also consistent with similar evidence from the SSD literature (Lewis et al., 2011; Shriberg et al., 2009; Tkach et al., 2011) and suggest that children with cognitive-linguistic SSD may have deficits specific to phonological capacity but not to mental manipulation, the second component in working memory.

An impairment in phonological storage does not necessarily support this study's original hypothesis. It was predicted that children with cognitive-linguistic SSD would have deficits in working memory that interfered with the ability to temporarily store and manipulate word forms undergoing correction. There was no evidence from the tasks

selected in this study to suggest that the clinical participants had difficulty with mental operations, and thus the proposed theoretical link between speech sound accuracy and working memory was not supported.

A deficit in phonological storage may have significant effects on earlier phonological development rather than transitioning to new word production form *per se*. For instance, phonological capacity could impact processes involved in learning native phonology (Stoel-Gammon, 2012). Similar to what has been proposed for SLI and word learning (Gathercole, 2006), the short-term memory store is theoretically involved in the process of forming stable phonological representations or native speech sound categories over time (Munson, Edwards & Beckman, 2012). Cowan and colleagues' model of working memory views storage as essentially synonymous with attention (Cowan, 2010); if the focus of attention is not adequately maintained or incorrect acoustic cues become the focus of attention, then underlying representations may not develop correctly. The present results provide further evidence for impaired phonological storage in children with cognitive-linguistic SSD, an area that deserves closer examination because of both theoretical and practical implications.

The fact that only one of three phonological working memory tasks was related to speech sound accuracy might have been due to extraneous factors such as task difficulty or age. As discussed in the results section, backward digit span proved to be so difficult that up to a quarter of participants were unable to perform the task during instructed practice. The animal span task proved to be very long and tedious, particularly for younger participants, a fact that was supported by a positive Pearson's correlation between task performance and age in months ($r = .311, p = .012$). The influence of

extraneous variables on performance again emphasizes the importance of task selection in this line of research (Carlson, 2005).

Organization of executive functions in preschoolers

Findings from this study also potentially contribute to the EF literature more broadly. Research in the preschool population has supported a number of different models of core EF organization during development, from a single or unitary EF construct to three distinct constructs similar to adult models (Diamond, 2013; Garon et al., 2008; Miller et al., 2012; Shing et al., 2010; Wiebe et al., 2008). Data analysis from the participants in this study instead supported a dual construct system of organization. As was evident from the tasks which factored together, children often accomplished tasks through reliance on alternate cognitive skills rather than those that were targeted (Dauveir et al., 2012; Ramscar et al., 2013), examples of which will be provided below.

The first group of tasks identified in the factor analysis consisted of forward and backward digit spans and the FIST. These results suggest that, for this age group, working memory may underlie these three tasks. While the digit spans tasks are intuitively obvious, successful performance on the FIST requires accurately storing the first selection while choosing the second target response. Both the FIST and forward digit span showed between-group differences and were strongly correlated. Notably, the FIST did not correlate with backward digit span, suggesting that only the storage component of working memory is shared by all three tasks. It would be interesting to investigate this relationship more closely as it relates to findings by Dodd and colleagues showing lower performance on the FIST by children with consistent but atypical errors (Crosbie et al.,

2009). These findings may reflect impaired phonological capacity in this population, rather than what has been proposed as a deficit in rule abstraction or cognitive flexibility.

The second group of tasks that factored together under a different latent variable included the Hearts and Flowers tasks, animal span, and the modified DNS. It seems likely that inhibitory control was the EF construct underlying these tasks. Both the Flowers, and Hearts and flowers tasks require cognitive control to inhibit the congruent or previous response. Animal span, although it was selected as a working memory task, likely required inhibition due to interference from previous trials. The modified DNS, which was specifically designed to tap into inhibitory control, was perhaps surprisingly the least weighted task in this construct and did not correlate with any other task. This result is consistent with Pasalich et al. (2010) who also found that, contrary to predictions, the DNS was not associated with performance on the more traditional preschool Stroop task. These findings appear to question the validity of the DNS as an inhibitory control task.

The present results in support of a dual construct model lead to a number of questions about cognitive flexibility and task-switching in children. It is possible that the tasks used in this study did not provide an adequate measure of this cognitive process. Card sorting tasks have been used in the adult literature to tap into this domain, and perhaps the inclusion of a child-friendly version would have yielded different results in the factor analysis. On the other hand, the ability to switch between tasks may be so reliant on working memory and inhibitory control abilities in preschoolers that cognitive flexibility should not be considered a separate construct until later in development.

Assessment of speech perception

This study explored the use of the SAILS program (Rvachew, 2010) to obtain a global measure of speech sound discrimination in children. Results demonstrated no significant difference between groups and no relationship between the speech perception scores and speech sound accuracy, particularly when age was partialled out. Because of these findings, a speech perception variable was not used in other analyses.

There are several likely explanations as to why this measure did not prove reliable. First, d-prime scores across five phonemes were used to obtain a global measure of speech sound discrimination. Although this composite method has been used effectively as a measure of speech sound consistency (Tyler et al., 2003), it may not be justified in speech perception (see also Munson et al., 2005). The SAILS program is designed to assess and treat specific phonemes that the child produces in error. Although four of the five phonemes selected for the measure were from the late-8 group in acquisition (Shriberg, 1993), speech sound discrimination at the preschool age may be phoneme-specific rather than a general deficit assessed by a single test (Rvachew et al., 2004).

Another possible explanation as to why this tool did not reveal differences between groups is that not all children with cognitive-linguistic SSD have been shown to have deficits in speech perception (Rvachew & Jamieson, 1989). An effect in a small group of children with these deficits could have been washed out in the group analysis. However, this explanation is less likely considering the lack of correlation between speech perception and speech sound accuracy, which would likely have revealed a relationship had there been one.

The final proposed explanation is more relevant to the task itself. As described in the methodology in the current study, the experimenter used a variety of ways to frame task instructions (e.g., right/wrong vs. good/bad) because many of the participants struggled to understand what was being asked of them. Even though all participants met the criterion on practice items, these trials appeared conceptually easier than experimental trials, in which differences between tokens were much more subtle. Although d-prime values were used to correct for response bias, scores may not have reflected children's actual abilities if the task was not understood. In this regard, SAILS may be a better treatment adjunct than experimental tool.

Non-word repetition and phonological storage

Several findings in this study contribute to evidence that children with cognitive-linguistic SSD have deficits in phonological storage capacity. Overall, children with SSD demonstrated significantly poorer performance in non-word repetition accuracy as compared to the typically-developing group. Likewise, accurate repetition of SRT items was positively correlated with speech sound accuracy. Children in the clinical group showed a similar length effect as their typically-developing peers, which further implicates phonological capacity wherein longer stimuli are more subject to trace decay than shorter items (Bowey, 2006; Graf Estes et al., 2007; Munson et al., 2005; Repovs & Baddeley, 2006). In addition, the positive correlation between digit span and non-word repetition was consistent with evidence from typically-developing children and other clinical populations demonstrating a similar relationship (see Gathercole, 2006, for review).

One concern in this study was that children with cognitive-linguistic SSD had both lower receptive vocabularies and lexical diversity in spontaneous speech than children in the typically-developing group. From a theoretical standpoint, a between-group difference in receptive and expressive vocabularies would be unlikely to significantly affect EF performance, but it could be important with regard to non-word repetition. Previous evidence has shown a strong relationship between vocabulary size and non-word repetition accuracy, which has been interpreted as either lexical knowledge capable of boosting non-word repetition performance or phonological memory facilitating both (Gathercole, 2006; Munson et al., 2012).

When receptive vocabulary was statistically controlled, there was no impact on the between-group difference on the SRT or in the correlation between SRT and speech production accuracy. Evidence suggests that the relationship between receptive vocabulary and non-word repetition significantly decreases with age starting around three or four years of age (Melby-Lervag et al., 2012; Gathercole, 2006), which might explain why controlling receptive vocabulary for these participants did not affect non-word performance. It should also be noted that receptive vocabularies for all participants were above the 33rd percentile as required by the study's eligibility criteria, with a mean above the 70th percentile.

On the other hand, expressive vocabulary, as measured by lexical diversity in spontaneous speech, not only affected the between-group significance in mod-SRT performance when entered as a covariate, but also showed a strong correlation to NWR accuracy. This study adds to the evidence base indicating lower standardized test scores in expressive vocabulary and lexical diversity in spontaneous speech in children with

cognitive-linguistic SSD as compared to TD children (Edwards, Fox & Rogers, 2002; Shriberg et al., 2009). Researchers have proposed that phonology and lexical development work in tandem in the early years (Schwartz & Leonard, 1982; Stoel-Gammon, 2011). As the lexicon grows, underlying phonological representations are refined and form connections within a network (Storkel & Hoover, 2010). Children with cognitive-linguistic SSD appear to lag behind peers in this process, as is evident in both their phonological skills and productive vocabularies. In light of this proposed relationship, it seems reasonable to expect that lexical diversity would be associated with non-word repetition performance, which can be seen as an integral step of the word learning process.

A difference in lexical diversity could also be related to children's efficiency in phonological encoding during spontaneous speech rather than an impairment in expressive vocabulary knowledge. Stokes and colleagues (2013) recently proposed a relationship between non-word repetition and lexical diversity in terms of the dual-stream model of speech production (Hickok & Poeppel, 2007). In this model, speech processing can be accomplished via a dorsal route that pairs auditory input with the articulatory network or a ventral route that connects input with pre-existing lexicon and meaning. It is possible that children with cognitive-linguistic SSD who demonstrate both decreased accuracy in non-word repetition and lower lexical diversity in spontaneous speech have impaired dorsal pathways, whereby pre-stored phonological and articulatory templates are less accessible.

Non-word repetition and prosodic stress

One aspect of the original SRT that was explored was whether the presence of overt stress would affect task performance either for all participants or differentially across groups. Although it was reasoned that adding stress might make the original syllable strings more word-like, a factor that has been shown to improve performance on non-word repetition (e.g., Edward, Beckman & Munson, 2004), results showed no difference between stressed and equal stress conditions across participants or by group. These results were consistent with Shriberg et al.'s (2009) assertion that SRT items would be interpreted as having stress. Recent results by Archibald, Gathercole and Joanisse (2009) found that co-articulatory effects strongly influenced whether syllable strings were perceived as word-like (see also Archibald & Gathercole, 2007; Gupta et al., 2005). It is also likely that the falling intonation contour used with the presentation of original SRT items contributed to their word-likeness. These results indicate that the addition of overt lexical stress does not necessarily improve the sensitivity of the original SRT stimuli.

Adding the stressed condition, however, did enable two other analyses of interest. First, it was found that SRT items with the preferred trochaic stress pattern were actually produced less accurately by all participants than items with iambic stress, although the difference was not statistically significant. These findings are contrary to evidence in the literature showing a preference for the trochaic pattern in English (Gerken, 1994; Roy & Chiat, 2004). However, the results might have been confounded by the composition of stimuli. Specifically, the two 4-syllable items identified as trochaic were produced with dominant stress on the third syllable only (e.g., nədə'məbə), which is a less usual stress

pattern in English than the more common 'S-w-S-w pattern (Kehoe, 1997). In addition, there was one more iambic than trochaic item in the 3-syllable level. Although the analysis was based on proportions, these factors might have unfairly weighted the iambic pattern.

A more informative analysis examined whether syllable stress would affect accuracy. It was found that all participants produced consonants in unstressed syllables with poorer accuracy than consonants in stressed syllables. This finding is consistent with the developmental pattern of weak syllable omission (Gerken, 1994, 1998; Gupta, 2005; Roy & Chiat, 2004), although in this study, consonants in unstressed syllables were generally substituted rather than omitted. Although it has been observed that clinical populations often demonstrate disproportionate effects from linguistic factors (Roy & Chiat, 2004), children with cognitive-linguistic SSD were no more vulnerable to errors on unstressed consonants than the typically-developing group.

Non-word repetition, EFs, and cognitive-linguistic SSD

A unique contribution of this study was that the effect of executive functions on non-word repetition ability could be explored. It was found that only forward digit span was significantly correlated with performance on the mod-SRT, although animal span approached significance. As discussed previously, animal span was found to factor with inhibitory control tasks rather than working memory tasks most likely because of the interference between like-trials. It is possible that the mod-SRT also required inhibitory control for the same reason, since the stimuli consist of very few phonemes. Interweaving

the stressed SRT condition with the original equal stress items might actually have watered down the effect of interference between items by adding more variability.

The relationship between forward digit span, non-word repetition and speech sound accuracy was further explored using a regression model. This analysis was conducted to address published concerns over why some clinical populations repeat 2-syllable non-words - which require little phonological storage capacity - less accurately than typically-developing children, a prior finding that was consistent in this study as well (Graf Estes et al., 2007; Shriberg et al., 2009). Results from the regression indicated that non-word repetition and forward digit span shared some of the variance in speech sound accuracy, but also uniquely contributed to it.

The shared variance by these tasks could be attributed to phonological capacity, but the additional contribution from the SRT is likely a different cognitive process. These findings are largely in agreement with the phonological processing account of phonological memory (PPA; Bowey, 2006; Gathercole, 2006), which proposes the involvement of speech perception, phonological encoding and retrieval, and articulatory planning and production in addition to phonological storage capacity. Since the SRT largely controls for articulatory processes, the list of possible cognitive processes is further narrowed to speech sound perception and phonological encoding/retrieval that might be impaired in this population. After controlling for speech perception and articulation, Munson et al. (2005) interpreted their findings as suggestive of poor or under-specified underlying representations in children with cognitive-linguistic SSD. Though the exact cognitive processes underlying non-word repetition are yet to be

determined, this study adds to the evidence suggesting that both phonological storage and another cognitive process are most likely impaired in cognitive-linguistic SSD.

Limitations

Several limitations of this study were related to the composition of the clinical group. It is possible that a greater mean severity level of speech sound disorder would have resulted in significant group differences in more EF tasks. Similarly, a larger clinical group might have affected the results by adding statistical power. Another area of weakness was the effect of possible subgroups. No direct assessments of speech sound consistency or stimulability were used and the speech perception measure was not used in the EF or SRT analyses for reasons discussed previously; however, any of these factors could have influenced the results had they been controlled.

Clinical implications and future directions

Although the research hypotheses concerning the relationships between core executive functions and speech sound accuracy were not substantiated, several results from this study have important implications for this population. The results add to the evidence base suggesting that phonological memory and another cognitive component such as phonological encoding or quality of underlying representations are impaired in children with cognitive-linguistic SSD. Further insights into the effects of phonological neighborhoods, both probability and density in children with SSD (Munson et al., 2005; Storkel & Hoover, 2010), may help better define the nature of these cognitive processes.

A better understanding of these processes will likely lead to more targeted therapies in the future.

Another domain-general process that could be implicated in this disorder is implicit learning. Specifically, a failure in the implicit learning of native-phonology might explain why children with cognitive-linguistic SSD benefit from any intervention that makes phonological rules explicit for them. A deficit in implicit learning could have also partially accounted for lower performance on the FIST, which requires children to independently apply several sets of abstract rules. Theories of implicit learning have been used to explain features of SLI (e.g., Ullman & Pierpont, 2005), but have not been explored in atypical speech sound development.

The fact that phonological memory, as measured by digit span and non-word repetition, is impaired in both children with SLI and children with cognitive-linguistic SSD, raises the question of why these disorders lead to distinct symptomatology. That is, how does impaired phonological capacity lead to either problems acquiring syntax or native phonology? Further work in genetics may shed light on this issue, although it obviously complicates the use of phonological memory as a unique endophenotype for distinguishing either disorder. In terms of clinical utility, future research could explore use of the SRT and forward digit span to help distinguish preschoolers who require speech services from those who are more likely to transition to adult-like speech sound systems on their own.

Finally, findings from this study support research indicating that children with cognitive-linguistic SSD and typical language may lag behind peers on measures of expressive and/or receptive vocabularies and lexical diversity in spontaneous speech.

These findings have direct implications for clinical management. Although lexical deficits may be more subclinical in nature, this evidence might suggest that clinicians consider dual programming in which children are introduced to new vocabulary in addition to therapy to correct speech sounds.

Appendices

Appendix A. Picture Naming Task stimuli (PNT; Preston & Edwards, 2010)

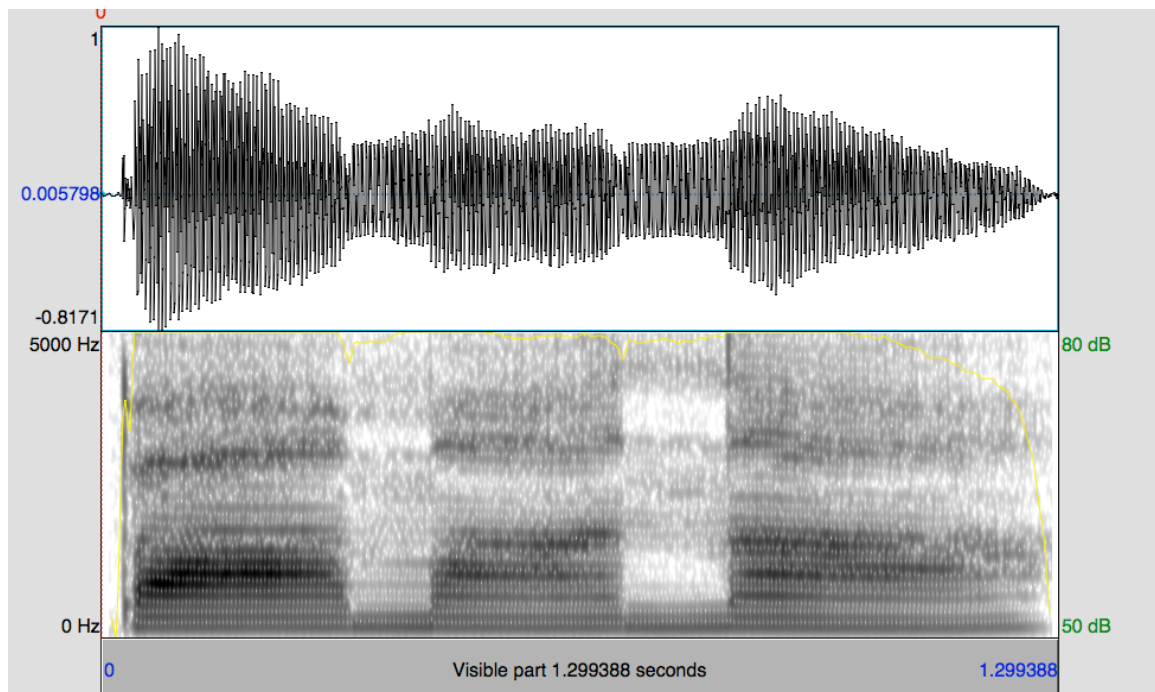
- | | | |
|---------------------|--------------------|--------------------|
| 1. parachute | 45. smooth | 89. yawn |
| 2. baby carriage | 46. snake | 90. elephant |
| 3. bathtub | 47. sleep | 91. plate |
| 4. beige | 48. swing | 92. present |
| 5. teeth | 49. splash | 93. blanket |
| 6. dinosaur | 50. spread | 94. breathe |
| 7. toy | 51. strawberry | 95. tree house |
| 8. ketchup | 52. screwdriver | 96. twins |
| 9. cookie | 53. squirrel | 97. pudding |
| 10. catch | 54. twelve | 98. dragon |
| 11. guitar | 55. queen | 99. crib |
| 12. measuring cup | 56. three | 100. quack |
| 13. newspaper | 57. skateboard | 101. glove |
| 14. giraffe | 58. ladybug | 102. green |
| 15. fire truck | 59. basket | 103. flower |
| 16. valentine | 60. chicken | 104. frog |
| 17. thimble | 61. pajamas | 105. throw |
| 18. this | 62. ice cream | 106. shrink |
| 19. scissors | 63. banana | 107. spider |
| 20. zebra | 64. telephone | 108. stamp |
| 21. xylophone | 65. television | 109. school bus |
| 22. shovel | 66. toothbrush | 110. smoke |
| 23. hippopotamus | 67. dishwasher | 111. snowman |
| 24. ladder | 68. cage | 112. slide |
| 25. refrigerator | 69. cowboy | 113. swimming pool |
| 26. washing machine | 70. garage | 114. splinter |
| 27. yoyo | 71. mailbox | 115. spring |
| 28. animals | 72. leaf | 116. string |
| 29. plant | 73. nose | 117. scratch |
| 30. princess | 74. chocolate | 118. squirtgun |
| 31. black | 75. jump rope | 119. clock |
| 32. brother | 76. jelly | 120. yellow |
| 33. bridge | 77. feather | 121. drum |
| 34. tractor | 78. vacuum cleaner | 122. dentist |
| 35. drive | 79. thank you | 123. washcloth |
| 36. clown | 80. thirsty | 124. hanger |
| 37. cracker | 81. there | 125. teacher |
| 38. glasses | 82. sandwich | |
| 39. grasshopper | 83. zipper | |
| 40. flag | 84. shampoo | |
| 41. french-fries | 85. helicopter | |
| 42. shrimp | 86. library | |
| 43. spaghetti | 87. rabbit | |
| 44. sticker | 88. window | |

Appendix B. Modified Syllable Repetition Task stimuli

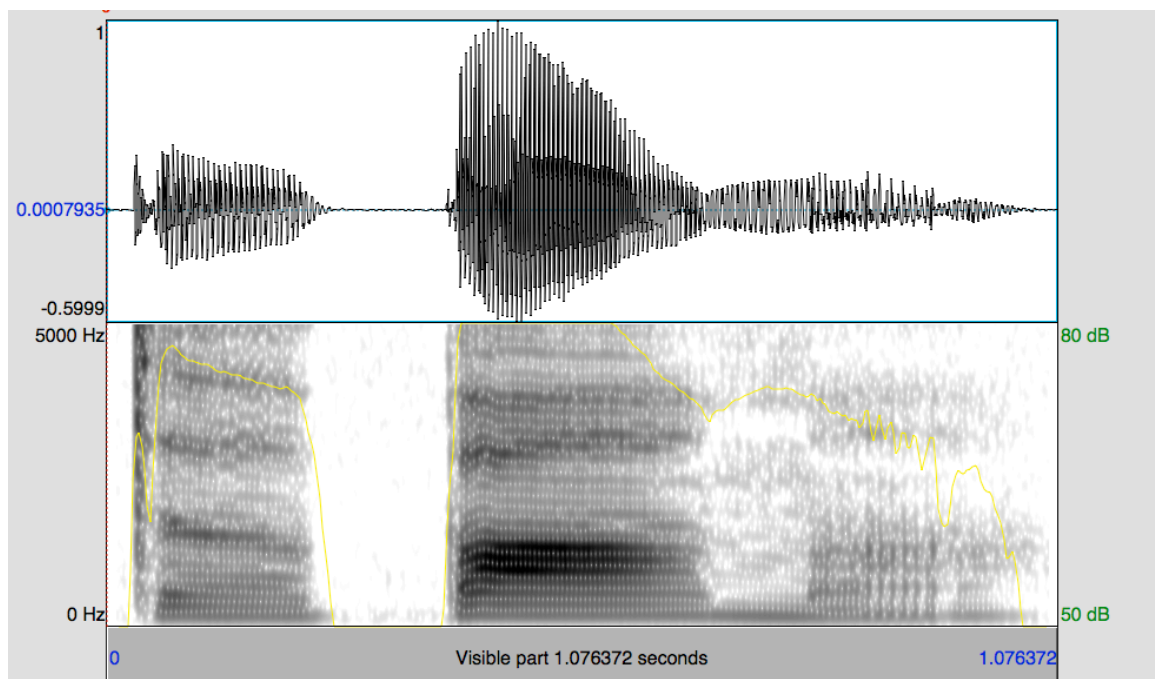
- | | |
|---------------------|------------------|
| 1. 'ba'da | 1. 'bamə |
| 2. 'ma'da | 2. nə'ba |
| 3. 'da'ma | 3. 'namə |
| 4. 'na'ba | 4. də'ba |
| 5. 'ba'ma'na | 5. də'bamə |
| 6. 'ma'da'ba | 6. 'nabədə |
| 7. 'da'ba'ma | 7. nə'damə |
| 8. 'ba'ma'da'na | 8. bə'namədə |
| 9. 'ma'na'ba'da | 9. nədə'mabə |
| 10. 'da'na'ba'ma | 10. məbə'danə |
| 11. 'na'da'ma'ba | 11. də'naməbə |
| 12. 'ma'ba'na'ma'da | 12. bə'namə,dabə |
| 13. 'na'da'ba'ma'na | 13. məbə'nadəmə |

Appendix C. Paired stimuli across conditions in modified SRT

Equal stress condition (original SRT)- /'ba'ma'na/



Stressed condition- /də'bamə/



References

- Adams, A. & Gathercole, S. (1996). Phonological working memory and spoken language development in young children. *The Quarterly Journal of Experimental Psychology*, 49, 216-233.
- Adams, A. & Gathercole, S. (2000). Limitations in working memory: Implications for language development. *International Journal of Language and Communication Disorders*, 35, 95-116.
- Aksan, N. & Kochanska, G. (2004). Links between systems of inhibition from infancy to preschool years. *Child Development*, 75, 1477-1490.
- Alloway, T., Gathercole, S., & Pickering, S. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development*, 77, 1698-1716.
- American National Standards Institute. (1991). *American national standard specifications for audiometers (ANSI S3.6-1969)*. New York, NY: ANSI.
- Amso, D. & Casey, B. (2006). Beyond what develops when: Neuroimaging may inform how cognition changes with development. *Current Directions in Psychological Science*, 15, 24-29.
- Archibald, L. & Gathercole, S. (2007). Nonword repetition in specific language impairment: More than a phonological short-term memory deficit. *Psychonomic Bulletin and Review*, 14, 919-924.
- Archibald, L., Gathercole, S., & Joanisse, M. (2009). Multisyllabic nonwords: More than a string of syllables. *Journal of the Acoustical Society of America*, 125, 1712-1722.

- Baddeley, A. (2001). Is working memory still working? *American Psychologist*, 56, 851-864.
- Badre, D. (2008). Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends in Cognitive Sciences*, 12, 193-200.
- Bernhardt, B. & Stoel-Gammon, C. (1994). Nonlinear phonology: Introduction and clinical application. *Journal of Speech and Hearing Research*, 37, 123-143.
- Bernthal, N., Bankson, J., & Flipsen, P. (2013). *Articulation and phonological disorders in children*, 7th edition. Boston, MA: Pearson Publishing.
- Biran, M. & Friedmann, N. (2005). From phonological paraphasias to the structure of the phonological output lexicon. *Language and Cognitive Processes*, 20, 589-616.
- Boersma, P. & Weenink, D. (2009). Praat: doing phonetics by computer, version 5.1.05. Retrieved November 1, 2011, from <http://www.praat.org>.
- Bowey, J. (2006). Clarifying the phonological processing account of nonword repetition. *Applied Psycholinguistics*, 27, 548-552.
- Briscoe, J & Rankin, P. (2009). Exploration of a 'double-jeopardy' hypothesis within working memory profiles for children with specific language impairment. *International Journal of Language and Communication Disorders*, 44, 236-250.
- Carlson, S. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology*, 28, 595-616.
- Chein, J. & Fiez, J. (2001). Dissociation of verbal working memory system components using a delayed serial recall task. *Cerebral Cortex*, 11, 1003-1014.

Chen, Z. & Cowan, N. (2005). Chunk limits and length limits in immediate recall: A reconciliation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 1235-1249.

Chevalier, N., Huber, K., Wiebe, S., & Andrews Espy, K. (2013). Qualitative change in executive control during childhood and adulthood. *Cognition*, 128, 1-12.

Chevalier, N., Sheffield, T., Nelson, J., Clark, C., Wiebe, S., & Andrews Espy, K. (2012). Underpinnings of the costs of flexibility in preschool children: The roles of inhibition and working memory. *Developmental Neuropsychology*, 37, 99-118.

Coady, J., & Evans, J. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *International Journal of Language and Communicative Disorders*, 43, 1-40.

Cocchi, L., Zalesky, A., Fornito, A., & Mattingley, J. (2013). Dynamic cooperation and competition between brain systems during cognitive control. *Trends in Cognitive Sciences*, 17, 493-501.

Cohen, J., MacWhinney, B., Flatt, M., and Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, 25, 257-271.

Cowan, N. (2010). Multiple concurrent thoughts: The meaning of developmental neuropsychology of working memory. *Developmental Neuropsychology*, 35, 447-474.

Crosbie, S., Holm, A. & Dodd, B. (2009). Cognitive flexibility in children with and without speech disorder. *Child Language Teaching and Therapy*, 25, 250-270.

Cummings, A. & Barlow, J. (2011). A comparison of word lexicality in the treatment of speech sound disorders. *Clinical Linguistics and Phonetics*, 25, 265-286.

Dauvier, B., Chevalier, N., & Blaye, A. (2012). Using finite mixture of GLMs to explore variability in children's flexibility in a task-switching paradigm. *Cognitive Development*, 27, 440-454.

Davidson, M., Amso, D., Cruess Anderson, L. & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, 44, 2037-2078.

Deak, G., Ray, S. & Pick, A. (2004). Effects of age, reminders and task difficulty on young children's rule-switching flexibility. *Cognitive Development*, 19, 385-400.

Dean, E. & Howell, J. (1986). Developing linguistic awareness: A theoretically based approach to phonological disorders. *British Journal of Disorders of Communication*, 21, 223-238.

Diamond, A. (2011). Biological and social influences on cognitive control processes dependent on prefrontal cortex. *Progress in Brain Research*, 89, 317-337.

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, 64, 135-168.

Diamond, A., Kirkham, N. & Amso, D. (2002). Conditions under which young children can hold two rules in mind and inhibit a prepotent response. *Developmental Psychology*, 38, 352-362.

Diamond, A. & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, 333, 959-964.

Diamond, A. & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and to "Do as I say, not as I do." *Developmental Psychobiology*, 29, 315-334.

Dockree, P. & Robertson, I. (2011). Electrophysiological markers of cognitive deficits in traumatic brain injury: A review. *International Journal of Psychophysiology*, 82, 53-60.

Dodd, B. (2005). *Differential diagnosis and treatment of children with speech disorder*, 2nd edition. London, England: Whurr Publishers.

Dodd, B. (2011). Differentiating speech delay from disorder: Does it matter? *Topics in Language Disorders*, 31, 96-111.

Dodd, B. & McIntosh, B. (2008). The input processing cognitive linguistic and oro-motor skills of children with speech difficulty. *International Journal of Language and Communication Disorders*, 10, 169-178.

Dollaghan, C. & Campbell, T. (1998). Nonword repetition and child language impairment. *Journal of Speech and Hearing Research*, 41, 1136-1146.

Dowsett, S. & Livesey, D. (2000). The development of inhibitory control in preschool children: Effects of "executive skills" training. *Developmental Psychobiology*, 36, 161-174.

Dunn, L. & Dunn, D. (2007). *Peabody Picture Vocabulary Test*, 4th edition. Circle Pines, MN: American Guidance Service Publishing/Pearson Assessments.

Edwards, J., Beckman, M., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research, 47*, 421-436.

Edwards, J., Fox, R., & Rogers, C. (2002). Final consonant discrimination in children: Effects of phonological disorder, vocabulary size, and articulatory accuracy. *Journal of Speech, Language, and Hearing Research, 45*, 231-242.

Farquharson Schussler, K. (2013). *Working memory processes in children with and without persistent speech sound disorders* (Unpublished doctoral dissertation). University of Nebraska: Lincoln, NE.

Fatzer, S. & Roebers, C. (2012). Language and executive functions: The effect of articulatory suppression on executive functioning in children. *Journal of Cognition and Development, 13*, 454-472.

Felsenfeld, S. & Broen, P. (1992). A 28-year follow-up of adults with a history of moderate phonological disorders: linguistic and personality results. *Journal of Speech and Hearing Research, 35*, 1114-1125.

Frank, M. (2006). Hold your horses: A dynamic computational role for the subthalamic nucleus in decision-making. *Neural Networks, 19*, 1120-1136.

Garon, N., Bryson, S., & Smith, I. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*, 31-60.

Gathercole, S. (1995). Nonword repetition: More than just a phonological output task. *Cognitive Neuropsychology, 12*, 857-861.

- Gathercole, S. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, 27, 513-543.
- Gerber, A. (1973). *Goal: Carryover*. Philadelphia: Temple University.
- Gerken, L. (1994). A metrical template account of children's weak syllable omissions. *Journal of Child Language*, 21, 565-584.
- Gerken, L. & McGregor, K. (1998). An overview of prosody and its role in normal and disordered child language. *American Journal of Speech-Language Pathology*, 7, 38-48.
- Gerstadt, C., Hong, Y. & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3 ½- 7 years old on a Stroop-like day-night test. *Cognition*, 53, 129-153.
- Gierut, J. (1991). Homonymy in phonological change. *Clinical Linguistics and Phonetics*, 5, 119-137.
- Gierut, J. (1998). Treatment efficacy: Functional phonological disorders in children. *Journal of Speech, Language and Hearing Research*, 41, 85-100.
- Gierut, J., Morrisette, M. & Ziemer, S. (2010). Nonwords and generalization in children with phonological disorders. *American Journal of Speech-Language Pathology*, 19, 167-177.
- Goldman, R. & Fristoe, M. (2000). *Goldman-Fristoe test of articulation*. Minnesota: American Guidance Services, Inc.
- Graf Estes, K., Evans, J., & Else-Quest, N. (2007). Differences in the nonword repetition performance of children with and without specific language

impairment: A meta-analysis. *Journal of Speech, Language and Hearing Research*, 50, 177-195.

Gruber, F. (1999). Probability estimates and paths to consonant normalization in children with speech delay. *Journal of Speech, Language and Hearing Research*, 42, 448-459.

Gupta, P., Lipinski, J., Abbs, B., & Lin, P. (2005). Serial position effects in nonword repetition. *Journal of Memory and Language*, 53, 141-162.

Hardcastle, W., Gibbon, F. & Jones, W. (1991). Visual display of tongue-palate contact: Electropalatography in the assessment and remediation of speech disorders. *International Journal of Language and Communication Disorders*, 26, 41-74.

Hasketh, A. (2010). Metaphonological intervention: Phonological awareness therapy. In A.L. Williams, S. McLeod & R.J. McCauley (Eds.), *Interventions for Speech Sound Disorders* (pp. 247-274). Baltimore, MD: Paul H. Brooks Publishing Co.

Hickok, G. & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews: Neuroscience*, 8, 393-402.

Hoff, E., Core, C. & Bridges, K. (2008). Non-word repetition assesses phonological memory and is related to vocabulary development in 20- to 24-month-olds. *Journal of Child Language*, 35, 903-016.

Hulterstam, I. & Nettelbladt, U. (2002). Clinician elicitation strategies and child participation: Comparing two methods of phonological intervention. *Logopedics Phoniatics Vocology*, 27, 155-168.

- Jacques, S. & Zelazo, P. (2001). The Flexible Item Selection Task (FIST): A measure of executive function in preschoolers. *Developmental Neuropsychology*, 20, 573-591.
- Jaeggi, S., Buschkuhl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training. *PNAS*, 108, 10081-10086.
- Johnson, S. & Somers, H. (1978). Spontaneous and imitated responses in articulation testing. *International Journal of Communication Disorders*, 13, 107-116.
- Kaufman, A. & Kaufman, N. (2012). *Kaufman Brief Intelligence Test*, 2nd edition. San Antonio, TX: Pearson Education.
- Keating, P. (2005). D-prime: Signal detection analysis. Retrieved from www.linguistics.ucla.edu/faciliti/facilities/statistics/dprime.htm
- Kehoe, M. (1997). Stress error patterns in English-speaking children's word productions. *Clinical Linguistics and Phonetics*, 11, 389-409.
- Kirkham, N., Cruess, L. & Diamond, A. (2003). Helping children apply to your knowledge to their behavior on a dimension-switching task. *Developmental Science*, 6, 449-476.
- Klein, E. (1996). Phonological/traditional approaches to articulation therapy: A retrospective group comparison. *Language, Speech and Hearing Services in Schools*, 27, 314-323.
- Koch, I., Gade, M., Schuch, S. & Philipp, A. (2010). The role of inhibition in task switching: A review. *Psychological Bulletin and Review*, 17, 1-14.
- Kornfeld, J. & Goehl, H. (1974). A new twist to an old observation: Kids know more than they say. Chicago, IL: Chicago Linguistics Society.

Law, J., Garrett, Z. & Nye, C. (2010). Speech and language therapy interventions for children with primary speech and language delay or disorder (Review). The Cochrane Collaboration: Wiley Publishers.

Lewis, B., Avrich, A., Freebairn, L., Taylor, H., Iyengar, S. & Stein, C. (2011). Subtyping children with speech sound disorders by endophenotypes. *Topics in Language Disorders, 31*, 112-127.

Lewis, B., Freebairn, L. & Taylor, H. (2000). Academic outcomes in children with histories of speech sound disorders. *Journal of Communication Disorders, 33*, 11-30.

Lewis, B., Shriberg, L., Freebairn, L., Hansen, A., Stein, C., Taylor, H., Iyengar, S. (2006). The genetic bases of speech sound disorders: Evidence from spoken and written language. *Journal of Speech, Language, and Hearing Research, 49*, 1294–1312.

Lof, G. (1996). Factors associated with speech-sound stimulability. *Journal of Communication Disorders, 29*, 255-278.

Low, J. & Simpson, S. (2012). Effects of labeling on preschoolers' explicit false belief performance: Outcomes of cognitive flexibility or inhibitory control? *Child Development, 83*, 1072-1084.

MacRoy-Higgins & Schwartz, R. (2013). Phonology and the lexicon in late talkers. L. Rescorla & P. Dale (Eds.), *Late Talkers: Language development, interventions, and outcomes* (pp. 113-128). Baltimore, MD: Brookes.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. Third Edition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Marcovitch, S. & Zelazo, P. (2009). A hierarchical competing systems model of the emergence and early development of executive function. *Developmental Science*, 12, 1-25.

Marton, K. (2006). Do nonword repetition errors in children with specific language impairment reflect a weakness in an unidentified skill specific to nonword repetition or a deficit in simultaneous processing? *Applied Psycholinguistics*, 27, 569-573.

McCormack, J., McLeod, S., McAllister, L. & Harrison, L. (2009). A systematic review of the associations between childhood speech impairment and participation across the lifespan. *International Journal of Speech-Language Pathology*, 11, 155-170.

McCormack, J., McLeod, S., McAllister, L. & Harrison, L. (2010). My speech problem, your listening problem, and my frustration: The experience of living with childhood speech impairment. *Language, Speech, and Hearing Services in the Schools*, 41, 379-392.

Melby-Lervag, M., Lervag, A., Halaas, S., Lyster, H., Klem, M., Hagtvet, B., & Hulme, C. (2012). Nonword-repetition ability does not appear to be a causal influence on children's vocabulary development. *Psychological Science*, 23, 1092-1098.

Miccio, A., Elbert, M., & Forrest, K. (1999). The relationship between stimulability and phonological acquisition in children with normally developing and disordered phonologies. *American Journal of Speech-Language Pathology*, 8, 347-363.

Miller, M., Giesbrecht, G., Muller, U., McInerney, R., & Kerns, K., (2012). A latent variable approach to determining the structure of executive function in preschool children. *Journal of Cognition and Development, 13*, 395-423.

Miyake, A., Friedman, N., Emerson, M., Witzki, A. & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49-100.

Morgan, J., Edwards, S., & Wheeldon, L. (2013). The relationship between language production and verbal short-term memory: The role of stress grouping. *The Quarterly Journal of Experimental Psychology, 67*, 220-246.

Morrisette, M. & Gierut, J. (2002). Lexical organization and phonological change in treatment. *Journal of Speech, Language, and Hearing Research, 45*, 143-159.

Munson, B., Edwards, J. & Beckman, M. (2005). Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. *Journal of Speech, Language and Hearing Research, 48*, 61-78.

Munson, B., Edwards, J. & Beckman, M. (2012). Phonological representations in language acquisition: Climbing the ladder of abstraction. In A.C. Cohn, C. Fougeron & M.K. Huffman (Eds.), *The Oxford Handbook of Laboratory Phonology* (288-309). Oxford, NY: Oxford University Press.

Nutley, S., Soderqvist, S., Bryde, S., Thorell, L., Humphreys, K. & Klingberg, T. (2011). Gains in fluid intelligence after training non-verbal reasoning in 4-year-old children: A controlled, randomized study. *Developmental Science, 14*, 591-601.

Østby, Y., Tamnes, C., Fjell, A., & Walhovd, K. (2011). Morphometry and connectivity of the fronto-parietal verbal working memory network in development. *Neuropsychologia*, 49, 3854-3862.

O'Sullivan, Mitchell, L. & Daehler, M. (2001). Representation and perseveration: Influences on young children's representational insight. *Journal of Cognitive and Development*, 2, 339-365.

Pasalich, D., Livesey, D., & Livesey, E. (2010). Performance on Stroop-like assessments of inhibitory control by 4- and 5-year-old children. *Infant and Child Development*, 19, 252-263.

Paynter, E. & Bumpas, T. (1977). Imitative and spontaneous articulatory assessment of three-year-old children. *The Journal of Speech and Hearing Disorders*, 42, 119-125.

Powell, T. (2008). The use of nonspeech oral motor treatments for developmental speech sound production disorders: interventions and interactions. *Language, Speech, and Hearing Services in Schools*, 39, 374-379.

Preston, J. & Edwards, M., (2007). Phonological processing skills of adolescents with residual speech sound errors. *Language, Speech, and Hearing Services in Schools*, 38, 297-308.

Preston, J. & Edwards, M. (2010). Phonological awareness and types of sound errors in preschoolers with speech sound disorders. *Journal of Speech, Language and Hearing Research*, 53, 44-60.

- Preston, J., Molfese, P., Mencl, W., Frost, S., Hoeft, F., Fulbright, R., . . . Pugh, K. (2014). Structural brain differences in school-age children with residual speech sound errors. *Brain and Language, 128*, 25-33.
- Ramscar, M., Dye, M., Gustafson, J., & Klein, J. (2013). Dual routes to cognitive-flexibility: Learning and response-conflict resolution in the dimensional change card sort task. *Child Development, 84*, 1308-1323.
- Repovs, G. & Baddeley, A. (2006). The multi-component model of working memory: Explorations in experimental cognitive psychology. *Neuroscience, 139*, 5-21.
- Rescorla, L. & Bernstein Ratner, N. (1996). Phonetic profiles of typically developing and language-delayed toddlers. *Journal of Speech and Hearing Research, 39*, 153-165.
- Robbins, J. & Klee, T. (1987). Clinical assessment of oropharyngeal motor development in young children. *Journal of Speech and Hearing Disorders, 52*, 271-277.
- Rose, Y., MacWhinney, B., Byrne, R., Hedlund, G., Maddocks, K., O'Brien, P., & Wareham, T. (2006). Introducing Phon: A software solution for the study of phonological acquisition. In D. Bamman, T. Magnitskaia & C. Zaller (Eds.), *Proceedings of the 30th Annual Boston University Conference on Language Development*, 489-500. Somerville, MA: Cascadilla Press.
- Roy, P. & Chiat, S. (2004). A prosodically controlled word and nonword repetition task for 2- to 4-year-olds: Evidence from typically-developing children. *Journal of Speech, Language and Hearing Research, 47*, 223-234.

Rvachew, S. (2010). The Speech Assessment and Interactive Learning System (SAILS). Retrieved November 20, 2012, from www.flintbox.com.

Rvachew, S. & Brosseau-Lapr , F. (2012). *Developmental phonological disorders: Foundations of clinical practice*. San Diego, CA: Plural Publishing.

Rvachew, S., Chiang, P., & Evans, N. (2007). Characteristics of speech errors produced by children with and without delayed phonological awareness skills. *Language, Speech and Hearing Services in the Schools*, 38, 60-71.

Rvachew, S. & Jamieson, D.G. (1989). Perception of voiceless fricatives by children with a functional articulation disorder. *Journal of Speech and Hearing Disorders*, 54, 193-208.

Rvachew, S., Nowak, M. & Cloutier, G. (2004). Effect of phonemic perception training on speech production and phonological awareness skills of children with expressive phonological delay. *American Journal of Speech-Language Pathology*, 13, 250-263.

Rvachew, S., Rafaat, S. & Martin, M. (1999). Stimulability, speech perception skills, and the treatment of phonological disorders. *American Journal of Speech-Language Pathology*, 8, 33-43.

Schoemaker, K., Bunte, T., Wiebe, S., Andrews Espy, K., Dekovic, M. & Matthys, W. (2012). Executive function deficits in preschool children with ADHD and DBD. *Journal of Child Psychology and Psychiatry*, 53, 111-119.

Schwartz, R. & Leonard, L. (1982). Do children pick and choose? An examination of phonological selection and avoidance in early lexical acquisition. *Journal of Child Language*, 9, 319-336.

Seeff-Gabriel, B., Chiat, S., & Dodd, B. (2010). Sentence imitation as a tool in identifying expressive morphosyntactic difficulties in children with severe speech difficulties. *International Journal of Language and Communication Disorders*, 45, 691-702.

Shipstead, Z., Redick, T., & Engle, R. (2012). Is working memory training effective? *Psychological Bulletin*, 138, 628-654.

Shing, Y., Lindenberger, U., Diamond, A., Li, S., & Davidson, M. (2010). Memory maintenance and inhibitory control differentiate from early childhood to adolescence. *Developmental Neuropsychology*, 35, 679-697.

Shriberg, L. (1993). Four new speech and prosody-voice measures for genetics research and other studies in developmental phonological disorders. *Journal of Speech and Hearing Research*, 36, 105-140.

Shriberg, L., Austin, D., Lewis, B., McSweeney, J. & Wilson, D. (1997). The Percentage of Consonant Correct (PCC) metric: Extensions and reliability data. *Journal of Speech, Language and Hearing Research*, 40, 708-722.

Shriberg, L., Fourakis, M., Hall, S., Karlsson, H., Lohmeier, H., McSweeney, J., . . . Wilson, D. (2010). Extension to the Speech Disorders Classification System (SDCS). *Clinical Linguistics and Phonetics*, 24, 795-824.

Shriberg, L. & Kwiatkowski, J. (1982). Phonological disorders III: A procedure for assessing severity of involvement. *Journal of Speech, Language and Hearing Disorders*, 47, 256-270.

Shriberg, L. & Kwiatkowski, J. (1994). Developmental phonological disorders I: A clinical profile. *Journal of Speech, Language and Hearing Research*, 37, 1100-1126.

Shriberg, L., Kwiatkowski, J., & Hoffman, K. (1984). A procedure for phonetic transcription by consensus. *Journal of Speech and Hearing Research*, 27, 456-465.

Shriberg, L., Lewis, B., Tomblin, J., McSweeney, J., Karlsson, H., & Scheer, A. (2005). Toward diagnostic and phenotype markers for genetically transmitted speech delay. *Journal of Speech, Language and Hearing Research*, 48, 834-852.

Shriberg, L. & Lof, G. (1991). Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics and Phonetics*, 5, 225-279.

Shriberg, L., & Lohmeier, H. (2008). *The Syllable Repetition Task (SRT)*. (Tech. Rep. No. 14). Phonology Project, Waisman Center, University of Wisconsin-Madison.

Shriberg, L., Lohmeier, H., Campbell, T., Dollaghan, C., Green, J., & Moore, C. (2009). A nonword repetition task for speakers with misarticulations: The Syllable Repetition Task (SRT). *Journal of Speech, Language and Hearing Research*, 52, 1189-1212.

Shuster, L. (1998). The perception of correctly and incorrectly produced /r/. *Journal of Speech, Language and Hearing Research*, 41, 941-950.

Simpson, A. & Riggs, K. (2007). Under what conditions do young children have difficulty inhibiting manual actions? *Developmental Psychology*, 43, 417-428.

- St Clair-Thompson, H. (2011). Executive functions and working memory behaviors in children with a poor working memory. *Learning and Individual Differences, 21*, 409-414.
- Stackhouse, J. & Wells, B. (1993). Psycholinguistic assessment of developmental speech disorders. *European Journal of Disorders of Communication, 28*, 331-348.
- Stein, C., Lu, Q., Elston, R., Freebairn, L., Hansen, A., Shriberg, L., . . . Iyengar, S. (2010). Heritability estimation for speech-sound traits with developmental trajectories. *Behavior Genetics, 41*, 184-191.
- Stoel-Gammon, C. (2011). Relationship between lexical and phonological development in young children. *Journal of Child Language, 38*, 1-34.
- Stokes, S. & Klee, T. (2009). The diagnostic accuracy of a new test of early nonword repetition for differentiating late talking and typically developing children. *Journal of Speech, Language and Hearing Research, 52*, 872-882.
- Stokes, S., Moran, C. & George, A. (2013). Nonword repetition and vocabulary use in toddlers. *Topics in Language Disorders, 33*, 224-237.
- Storkel, H. & Hoover, J. (2010). Word learning by children with phonological delays: Differentiating effects of phonotactic probability and neighborhood density. *Journal of Communication Disorders, 43*, 105-119.
- Strombergsson, S., Wengelin, A., & House, D. (2014). Children's perception of their synthetically corrected speech production. *Clinical Linguistics and Phonetics, Early Online*, 1-23.

- Stuss, D., Levine, B., Alexander, M., Hong, J., Palumbo, C., Hamer, L., . . . Izukawa, D. (2000). Wisconsin Card Sorting Test performance in patients with focal frontal and posterior brain damage: Effects of lesion location and test structure on separable cognitive processes. *Neuropsychologia*, 38, 388-402.
- Sutherland, D. & Gillon, G. (2007). Development of phonological representations and phonological awareness in children with speech impairment. *International Journal of Language and Communication Disorders*, 42, 229-250.
- Taylor, C., Christensen, D., Lawrence, D., Mitrou, F. & Zubrik, S. (2013). Risk factors for children's receptive vocabulary development from four to eight years in the longitudinal study of Australian children. *PLOS One*, 8, 1-20.
- Templin, M. (1947). Spontaneous versus imitated verbalization in testing articulation in preschool children. *Journal of Speech Disorders*, 12, 293-300.
- Thal, D., Oroz, M. & McCaw, V. (1995). Phonological and lexical development in normal and late-talking toddlers. *Applied Psycholinguistics*, 16, 407-424.
- Tkach, J., Chen, X., Freebairn, L., Schmithorst, V., Holland, S., & Lewis, B. (2011). Neural correlates of phonological processing in speech sound disorder: A functional magnetic resonance imaging study. *Brain and Language*, 119, 42-49.
- Tyler, A., Lewis, K. & Welch, C. (2003). Predictors of phonological change following intervention. *American Journal of Speech-Language Pathology*, 12, 289-298.
- Tyler, A., Williams, M. & Lewis, K. (2006). Error consistency and the evaluation of treatment outcomes. *Clinical Linguistics & Phonetics*, 20, 411-422.

- Ullman, M. & Pierpont, E. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*, 41, 399-433.
- Waring, R. & Knight, R. (2013). How should children with speech sound disorders be classified? A review and critical evaluation of current classification systems. *International Journal of Communication Disorders*, 48, 25-40.
- Wechsler, D. (1991). *Wechsler Adult Intelligence Scale- Third Edition*. San Antonio, TX: The Psychological Corporation.
- Wiebe, S., Andrews Espy, K. & Charak (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology*, 44, 575-587.
- Wiig, E., Secord, W. & Semel, E. (2004). *Clinical Evaluation of Language Fundamentals in Preschool* (2nd ed.). San Antonio, TX: The Psychological Corporation.
- Williams, A. (2003). *Speech disorders: A resource guide for preschool children*. Clifton Park, NY: Thompson Delmar Learning.
- Willoughby, M., Blair, C., Wirth, R., Greenburg, M. & The Family Life Project Investigators (2012). The measurement of executive function at age 5: Psychometric properties and relationship to academic achievement. *Psychological Assessment*, 24, 226-239.
- Winitz, H. (1969). *Articulatory Acquisition and Behavior*. New York, NY: Appleton-Century-Crofts.
- Wood, J. (2007). Understand and computing cohen's kappa: A tutorial. Retrieved from http://wpe.info/papers_table.html

Yerys, B. & Munakata, Y. (2006). When labels hurt but novelty helps: Children's perseveration and flexibility in a card-sorting task. *Child Development*, 77, 1589-1607.

Yerys, B., Wolff, B., Moody, E., Pennington, B., & Hepburn, S. (2012). Brief report: Impaired Flexible Item Selection Task (FIST) in school-age children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 42, 2013-2020.

Zelazo, P. (2004). The development of conscious control in childhood. *Trends in Cognitive Sciences*, 8, 12-17.

Zelazo, P., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, 11, 37-63.

Zhang, Y. & Francis, A. (2010). The weighting of vowel quality in native and non-native listeners' perception of English lexical stress. *Journal of Phonetics*, 38, 260-271.